Institute for Physical AI @ Bailey Military Institute · The Charlot Lab

SIM TO REAL · EMBODIED POLICY · PHYSICAL AGENCY

Technical Report TR-2026-35
Research / Review
Preprint v2
14 August 2026

SIM TO REAL · EMBODIED POLICY · PHYSICAL AGENCY

Sim to Real: The Road to Physical Agency

What separates a policy that works in simulation from one that works on a body, read in eleven languages against the Institute's own gaps.

Dean David Jean Charlot, Industrial Research Fellow · The Charlot Lab

The Charlot Lab. Evidence base assembled 8 to 14 August 2026. Version 2 supersedes v1; see Section 2.

11 language regions | 504 graded claims | 13 gaps re-adjudicated | 43 contradictions adjudicated

A policy that works in simulation and fails on a body has met one of four walls: the world model is wrong, the action-conditioned transition function is wrong, a sensing channel has degraded rather than failed, or the body itself has drifted. This review asks what the record measures at each wall, and it asks in eleven languages, because the English-language conversation is not the record. Three evidence bases are used and kept apart. A six-chapter practitioner session, 179 claims of which 123 grade self-published, since a conference talk is self-published by its speaker whatever the affiliation behind it. A first sweep of Chinese, Japanese, Korean, German and French, 179 findings, 70 verified. A second sweep of Russian, Hindi and Indian, Portuguese, Spanish, Italian and Nordic, 146 findings, which returned 43 claims of contradicting the first. Those were adjudicated: five are real, thirty are protocol differences that only look like conflicts, and eight are misreadings. That ratio is this version's most portable result. Most apparent disagreement in this field is two measurements of different things wearing the same name. Version 2 corrects two conclusions of version 1. The vocabulary map it reported is real for Japanese, Chinese and Russian and does not generalise: four of six new regions report no vocabulary gap, and one has an absence of the literature rather than a hidden one. And gap 11, the only gap version 1 closed, moves to partly closed,

> because the operative standard has moved container, the measurement method sits in a separate instrument, and a vendor tolerance of 25 N stands against a 150 N limit. Re-adjudicated across eleven regions, all thirteen gaps moved and seven changed constraint class, most of them away from technology and toward measurement. Joules per embodied task remains not located as a measured quantity in any of the eleven.

1. Scope, method and grading

This review has one question. When a policy that works in simulation is put on a body and fails, what exactly failed, and does anyone measure it. Practitioners in the session that opens this review name four walls, and they are used here as the organising axis. Wall 1 is physical world modelling: the geometry, the materials and the contact are wrong, so the policy is solving a different problem from the one in front of it. Wall 2 is the action-conditioned transition function: predicting the next state from the current state is tractable, and conditioning that prediction on the action is the rung that separates a spectator model from an agent. Wall 3 is policy robustness when a sensing channel degrades rather than fails. Wall 4 is embodiment drift, the state-action-to-force map changing as actuators take on dust, corrosion and wear. The four fail differently, they are measured differently, and a result that closes one says nothing about the others.

Two evidence bases are used and they are kept apart throughout. The first is a six-chapter practitioner session, from which 179 claims were extracted with their quantities. Of those, 123 grade self-published. That is not a criticism of the speakers. Under the Institute rule a conference talk is self-published by the party giving it, and six Stanford and Y Combinator affiliations do not change the grade. Any reader who takes the practitioner base as evidence of the field's state should hold that ratio in view.

The second and third bases are two sweeps of non-Anglophone literatures, conducted in each region's own language and venues rather than in English translation. The first covered Chinese, Japanese, Korean, German and French and returned 179 findings, 127 terms of art flagged as not translating, and 74 records of quantities searched for and not located. The second covered Russian, Hindi and Indian, Portuguese, Spanish, Italian and Nordic and returned 146 findings and 43 claims of contradicting the first. Across both, 37 load-bearing figures were sent to independent readers instructed to refute them and to default to refuted unless they

could open the source and see the value themselves. Eight did not survive and are excluded from every count here.

The two bases invert on grade, and the inversion is itself a finding. The practitioner base is 69 percent self-published. The non-Anglophone base is 70 verified against 54 reported, 38 self-published and 16 modelled. The non-English record is more heavily standards, certification and peer review, and less heavily conference stage. A survey that reads only English is not sampling the same population.

Four grades are used and every figure carries one. Verified means the primary document was retrieved and the number read from it. Reported means a named source states it and this review did not confirm it against the primary. Self-published means the claimant and the source are the same party, including that party's own forward-looking targets, whatever outlet relayed them. Modelled means arithmetic performed here on sourced inputs, which is never presented as a measurement.

One convention governs every absence in this document. Where it says a quantity was not located, that means these sweeps searched for it and did not find it, and the search is named. It does not mean the quantity is unpublished. Of the 74 not-located records, 28 failed on access rather than on absence, 11 of those on documents that are sold rather than posted. Treating an access failure as evidence about the world would be the same error this review exists to correct.

Evidence base	Claims	Verified	Self-published	What it is good for
Practitioner session, 6 chapters	179	22	123	Naming the open problems and the vocabulary practitioners use
Sweep one, five regions	179	70	38	Quantities, protocols and statutory instruments
Sweep two, six regions	146	see Section 2	see Section 2	Auditing sweep one, and four measured censuses
Adversarial verification	37 checked	29 survived	8 refuted	Removing figures that do not survive their own source

Table 1. The three evidence bases, kept apart because they grade differently. The practitioner base names problems; the two regional sweeps measure them, and the second audits the first.

2. Corrections and withdrawals, read first

This version supersedes v1, which reported a five-region sweep. A second sweep of six further regions, Russian, Hindi and Indian, Portuguese, Spanish, Italian and Nordic, returned 146 findings and 43 claims that it contradicted the first. Those 43 were adjudicated one by one against both primaries. Five are real contradictions, thirty are protocol differences that only look like conflicts, and eight are misreadings. The corrections below change conclusions this Institute published, and they are stated at the front rather than buried, because a reader of v1 who does not reach the back of v2 should still meet them.

The first correction is to this review's own headline. Version 1 reported a vocabulary map: terms of art in non-English literatures that do not translate and that an English-only search cannot reach. That result is real for Japanese, Chinese and Russian. It does not generalise. Asked the same question with instruction not to manufacture a gap, four of the six new regions reported none. Italian borrows the field's terms whole rather than translating them. Nordic quantitative results are published in English in venues already reachable. Hindi carries an extensive standardised terminology, roughly 2.2 million terms across 22 languages through a state commission, with no primary robotics research written in it. Brazil has an absence rather than a hidden literature: a hand count of 2,159 titles across four national proceedings returned zero papers on sim-to-real transfer, domain randomisation, imitation learning, learned grasping, friction identification, vision-language-action models or diffusion policies. The barrier is language in some places, venue in others, and in one place the work is simply not there. A single explanation was wrong.

The second correction is to the only gap version 1 closed. Gap 11, the human-biomechanical contact field, moves from closed to partly closed on three grounds. The operative container has moved: ISO 10218-2:2025 absorbs the collaborative-application requirements including the Annex A biomechanical table, and ISO/TS 15066:2016 remains catalogued and flagged for revision with ISO/AWI 15066-1 under development, so it is superseded in practice rather than withdrawn. Version 1 recorded the limits and not the measurement method, and ISO/PAS 5672:2023 specifies the measurement and analysis methods and the pressure-force device characteristics: a limit without a measurement method is not a closed field. And an

enforcement-layer error bar exists that version 1 did not hold, a declared performance-level tolerance of 25 N against a 150 N limit, verified across four language editions of one vendor manual.

The third correction is a measurement-discipline error that ran through version 1's tactile rows. Eight sensor figures were tabulated as if commensurable when they are quantisation claims with no measurement beside them. One Russian device publishes both for the same sensor: a 1 mN quoted resolution against a maximum radial deviation of 0.147 N, 147 times the resolution, and 5.1 percent mean relative error radially against 0.8 percent normally. Resolution is not accuracy, and the two are axis-dependent. Every sensor row in this review now carries its statistic type and bandwidth, and no ratio is taken across rows. The same error invalidated the arithmetic in two sweep-two claims, which are excluded.

Two quantities are withdrawn outright. A modelled sim-to-real tax of 8.7 to 11.7 percentage points against the hardest randomisation band, with its conservation-law gloss, does not survive: the underlying published figures stand and the derivation over them does not. And a widely relayed 89.4 percent in simulation against 12 percent in a real household is withdrawn as a quantity. Version 1 already recorded that it did not locate those figures in the primary document several outlets attribute them to; it is retained here only as an artifact of trade-press discourse, with no numeric standing. Fourteen further figures are qualified rather than withdrawn, each with the condition that makes it true stated beside it.

One methodological result deserves its own line, because it governs how the rest of this review should be read. Of 43 claimed contradictions, thirty were protocol differences. A distribution average over ten thousand rollouts across seventeen dynamics parameters was being compared with thirty rollouts at a single point; a visual randomisation ladder was being compared with a dynamics randomisation width; a success predicate was moving a result by 84 percentage points on the same ten rollouts. Most apparent disagreement in this field is not disagreement. It is two measurements of different things wearing the same name.

Correction	Version 1 said	Version 2 says	Ground
The vocabulary headline	Non-English literature carries untranslatable terms of art	True for Japanese, Chinese and Russian; four of six new regions report	Six regional assessments, instructed not to manufacture a gap

		no vocabulary gap	
Gap 11, contact field	closed	partly closed	Container moved to ISO 10218-2:2025; measurement method sits in ISO/PAS 5672:2023; a 25 N tolerance against a 150 N limit
Tactile sensor rows	Eight figures tabulated together	Not rankable; each carries statistic type and bandwidth	One device publishing 1 mN resolution against 0.147 N maximum deviation
Sim-to-real tax, modelled	8.7 to 11.7 pp against the hardest band	withdrawn; the source figures stand	The derivation does not survive the protocol difference
89.4 percent against 12 percent	carried as framing	withdrawn as a quantity	Not located in the primary in either sweep
43 claimed contradictions	not yet examined	5 real, 30 protocol, 8 misreading	Adjudicated against both primaries

Table 2. What changed between versions. The third column is what this review now holds; the fourth is why.

3. Four walls, and where the Institute stands against each

The four walls are not equally attended to, in the field or in this Institute. Across the six chapters, wall 1 and wall 4 draw five named open problems each, wall 3 draws three and wall 2 draws two. Across the Institute's own 25 relevant holdings the ordering is different again: wall 1 holds ten, wall 4 six, wall 3 five and wall 2 four. Both distributions put wall 2 last, and wall 2 is the one the practitioners describe as the rung that separates a model that watches from a model that acts.

Wall 1 is where the corpus is strongest and where the practitioner record is bluntest. The session states the sim-to-real gap as unsolved and offers no gap metric, no benchmark and no magnitude, which is a statement about the boundary

rather than about the difficulty. Deformable objects are named in one clause as strictly worse than the rigid case for both modelling and simulation, with no elaboration. Physics that current simulators do not represent is given by example rather than by class: water, a zipper, cracking open a can.

Wall 2 is stated precisely enough to be tested. Predicting the next state from the current state works; conditioning on the action is described as not working and as demanding far more data. Action-space representation is called an unsolved problem for anyone who wants to learn quickly, which frames it as sample efficiency rather than impossibility. Sections 5 and 6 take those two statements to the eleven-region record and ask what has been measured against them.

Wall 3 is the one with the cleanest thought experiment and the least evidence. The example given is a person with intact limbs and intact vision who cannot lace a skate with cold hands: the sensing channel is degraded rather than absent, and the policy fails. This review located no robot analogue, dataset or metric for that condition in any of the eleven regions, having searched for degraded-channel robustness in each. Wall 4, embodiment drift, is named with no wear rate, no time constant and no failure threshold anywhere in the session.

The bar the session sets for physical agency is worth recording because it names a capability rather than a rate: the backpack test, identifying the contents of a bag by feeling inside it, with no vision. That is non-visual world modelling from contact alone, and it sits across walls 1 and 3 at once. No system in this review attempts it. Where a paper needs a single sentence for what separates physical motion from physical agency, that is the sentence.

Wall	What fails	Open problems named in the session	Institute holdings
1 Physical world modelling	Geometry, materials and contact are wrong	5	10
2 Action-conditioned transition	Next state given the action is wrong	2	4
3 Degraded sensing channel	Channel degrades rather than fails; policy fails with it	3	5
		5	6

4 Embodiment drift	The state–action–to–force map changes with wear		

Table 3. The four walls, with attention counted on both sides. Wall 2 is last in both columns and is the one the session describes as separating a spectator model from an agent.

4. What an English-only search cannot reach

Two absences sit behind the English-language record, and they have different causes and different remedies. The first is material that exists in English, is indexed in English and is freely retrievable, yet carries no weight in the robot-learning conversation because it lives in standards bodies, certification regimes and industrial measurement practice. ISO/TS 15066:2016 Annex A publishes maximum permissible peak pressure at quasi-static contact for 29 body localisations, 130 N/cm² at the forehead centre and 110 N/cm² at the temple [46], and maximum transferable energy per body region from 0.11 J at the face to 2.6 J at the pelvis [49]. ISO 9283:1998 has carried drift of pose accuracy and drift of pose repeatability since 1998 [64]. RoboChallenge publishes a full real-robot grading rule with per-model results, pi0.5 at 43.7%, pi0 at 28.3% and CogACT at 11.7% [17]. All of that is English. The cost of this absence is citation practice, and the remedy is reading.

The second absence is material with no English-language existence at all, and it is the finding. A J-STAGE full-text search for Sim-to-Real together with 実機 returns 4,862 results, and 日本ロボット学会誌 alone held 7,006 papers as of 2026-07-31 [112]. Much of the highest-volume Japanese work appears there as four-page peer-reviewed レター rather than as full papers, fully open access. Inside that format sits バイラテラル制御に基づく模倣学習, in which demonstrations are collected on a leader-follower pair so that human-applied force and environment reaction force land on different robots and can be separated, after which the policy predicts force commands and runs force control. One letter reports per-step success on a six-step hamburger assembly across six initial layouts [86], trained on 48 demonstrations, seven for training and one for validation per layout, with a 500 Hz control loop decimated 20:1 to 25 Hz [87]. This review located no English-language name for the branch, having searched the robot-learning venues in English.

Contact force in that line is an observer output rather than an instrument reading. CRANE-X7 carries no torque sensor: disturbance torque is computed by 外乱オブザーバ and torque response is estimated by 反力推定オブザーバ, cited to work from 1996 and 1993 [88], on a 500 Hz control loop [87]. Shortening the four-channel bilateral

control period from 2 ms to 1 ms is reported to execute tasks up to 30% faster [91]. Adding a physical-consistency loss relating angle to angular velocity moved real-robot success from 82.2% to 94.4% on 15 demonstrations [90]. An English query on sensorless force control does not carry the two-observer decomposition, and it returns a different literature.

深層予測学習 is the Waseda school's name for learning an action-conditioned forward model and controlling through it. It predates and sits beside 世界モデル and does not translate as "world model" [111]. That bears on this Institute directly. Gap 1 of the thirteen gaps recorded against the corpus states that no holding takes the action-conditioned transition function as its subject, and part of what was recorded there as an absence of work was a failure of search vocabulary rather than an absence in the world. Adjacent Japanese holdings sit in the same untranslated record: stiffness emitted inside a predicted t+20 ms next state, discriminating 15 from 10 Nm/rad on five demonstrations per bottle [92], and a real-to-sim identification loop whose cost curve runs 2,989, 28,762 and 279,927 annealing trials at objective values 0.022276, 0.004156 and 0.000451 [94]. The gap narrows and does not close.

China's untranslated contribution is regulatory and industrial. YD/T 6770-2026 人工智能 关键基础技术 具身智能基准测试方法 was approved by MIIT in March 2026, entered force on 2026-06-01 and was drafted by CAICT with more than forty organisations, making five metrics reportable, among them 场景扰动衰减率 and 平均任务能耗 [21]. GB/T 40575-2021, an industrial-robot energy-efficiency instrument, was issued on 2021-10-11 [22], so energy entered Chinese standards infrastructure years before the embodied benchmark reached it. 数采 names an occupation and an industry segment rather than an activity: rigs above ¥200,000 and roughly ¥300 per collector-day against about 500 trajectories per collector-day [28], which prices labour at ¥60 to ¥150 per hour of demonstration data against a ¥500 to ¥1,000 per hour market price [29]. An English query on data collection returns method papers.

定格出力80W is a statutory line. Machines whose drive motor is rated at 80 W or less are excluded from the industrial-robot provisions of 労働安全衛生規則 第150条の4 [106], and the Japanese product market is organised around 「80W規制」. The neighbouring limits are stated in units that do not convert into it: reduced speed at the tool centre point shall not exceed 250 mm/s [107], and JIS B 8445:2016 excludes from its scope any 生活支援ロボット moving faster than 20 km/h [108]. A search on

collaborative safety in English returns none of this, because a watt rating at the actuator is not a category the English-language conversation holds.

Position on the trajectory for this Institute's own retrieval: 25 degrees of ninety. The binding constraint is workforce, meaning reading capacity in the source language, since nothing here is limited by compute, hardware or algorithm. The threshold quantity that moves it is coverage of the Japanese record: a standing sweep that issues 環境乱択化, ドメインランダマイゼーション and ドメインランダム化 as separate queries, since they return different paper sets, and that reaches the 4,862 J-STAGE records answering Sim-to-Real together with 実機 [112] rather than their English-abstract subset. This review did not locate a measurement of what fraction of those 4,862 records is retrievable by an English-only query, having searched J-STAGE in Japanese and in English.

Term in its own script	Gloss	What an English-only search fails to retrieve	Region
深層予測学習	deep predictive learning	The Waseda line on learning an action-conditioned forward model and controlling through it; it predates and sits beside 世界モデル, so an English query on world models does not reach it [111]	Japanese
バイラテラル制御に基づく模倣学習	imitation learning based on bilateral control	An imitation-learning branch published as four-page JRSJ letters, in which the policy predicts force commands; per-step success on a six-step assembly [86] from 48 demonstrations at 500 Hz decimated to 25 Hz [87]. No English name for it was located	Japanese
外乱オブザーバ / 反力推定オブザーバ	disturbance observer / reaction force estimation observer	The two-observer decomposition that recovers contact torque from motor current with no torque sensor in the arm [88], on a 500 Hz control loop [87]	Japanese

定格出力80W / 「80W規制」	rated output 80 W / the 80 W regulation	The statutory exemption of drive motors rated at 80 W or less from 労働安全衛生規則第150条の4 [106], the threshold around which the Japanese product market is organised	Japanese
YD/T 6770–2026	AI key foundational technologies: embodied intelligence benchmark test methods	The mandate of 平均任□能耗 and □景□□衰减率 as required reportables, in force 2026-06-01, drafted by CAICT with more than forty organisations [21]	Chinese
数采	demonstration-data production, as an occupation and an industry segment	Rigs above ¥200,000, roughly ¥300 per collector-day and about 500 trajectories per collector-day [28], pricing labour at ¥60 to ¥150 per hour of data [29]	Chinese
実機	the actual machine	The pole opposed to any model, analytical or learned, older than the reinforcement-learning literature; 4,862 J-STAGE records answer it together with Sim-to-Real [112]	Japanese
環境乱択化 / ドメインランダマイゼーション / ドメインランダム化	three current renderings of domain randomisation	Different paper sets per rendering, so a sweep issuing only one of them under-counts the corpus that reports the bands	Japanese

Terms whose own script is the only door, and what an English-only search fails to retrieve

5. The action-conditioned transition function, and action representation

EWMBench scores seven embodied world models on one protocol and separates appearance from dynamics. The protocol is stated in the same document: AgiBot-World data, 10 manipulation tasks by 10 ground-truth episodes, 4 to 10 atomic sub-

actions per task, 7 models by 3 candidates for 2,100 generated videos, with DINOv2 patch cosine similarity for scene consistency, symmetric Hausdorff distance and normalised dynamic time warping for motion, and Wasserstein distance on velocity and acceleration for dynamic consistency [24]. OpenSora scores 0.9210 on scene against the best model's 0.9427, and 0.0474 on dynamics against 0.5363 [24]. That is an 11.3x gap on dynamics at near parity on appearance. Appearance and action-conditioned dynamics are empirically decoupled, so a benchmark scoring pixels scores the wrong quantity. The competition-scale echo carries a weaker grade: the AGIBOT WORLD CHALLENGE ran 526 teams from 27 countries across three tracks, and its world-model track was won at 0.829 by a team that also ranked first on the motion-following sub-score, self-published because the organiser owns the benchmark and sells the hardware [25].

What the missing holding costs is measurable on real hardware. RoboChallenge's worst difficulty tag of nine is "temporal", 3 tasks at SR 5% and progress score 14, against an all-task mean of SR 22% and score 37 over 30 real tasks on four embodiments at 10 rollouts per task [16]. The report's own explanation is that every model tested is single-frame and therefore carries no transition function at all. GemBench Level 4 returns 0.0 ± 0.0 for Hiveformer, 3D Diffuser Actor and RVT-2, 0.1 ± 0.2 for PolarNet and 0.3 ± 0.3 for 3D-LOTUS, while the same methods score 60.3% to 94.3% at Level 1, and only 3D-LOTUS++ reaches 17.4 ± 0.4 [76]. Every entry there is a policy holding. All of it reads against a measured run-to-run variance of 0% to 100% with task, props and model held constant [15], which is larger than most of the effects being ranked.

Three Japanese letters take the transition itself as subject. 佐藤寛 et al., JRSJ 43(6):603-606 (2025), impose a physical-consistency loss stating that angle is the integral of angular velocity, moving real-robot success from 82.2% to 94.4% on 15 demonstrations for one task and 39 for the other, at 10 rollouts per distance including an untrained 24 cm [90]. JRSJ 44(4):433-436 (2026) emits stiffness inside the predicted t+20 ms next state, discriminating 15 from 10 Nm/rad on two 6 cm PET bottles from 5 demonstrations each [92]. JRSJ 44(5):524-527 (2026) relabels a hexapod command and recomputes the reward because the physical transition is invariant to the relabelling [109]. Part of what reads as absence is search vocabulary: 深層予測学習, deep predictive learning, does not translate as world model [111]. This review did not locate a holding stating a transition model's quality and its real-robot task effect under one protocol, having searched the Chinese, Japanese, Korean, German and French literatures in native language and native venues.

Discretisation against continuous action is closed in simulation with a protocol, and this review located no other closure of it. Garcia-Pinel holds backbone and conditioning identical and varies only the head under the stated GemBench protocol: regression with AdaptiveNorm gives L1 83.3 ± 0.7, L2 29.3 ± 1.9, L3 34.5 ± 1.0 and L4 0.0 ± 0.0; classification with AdaptiveNorm gives 90.8, 47.8, 37.9 and 0.0; classification with cross-attention gives 94.3 ± 1.4, 49.9 ± 2.2, 38.1 ± 1.1 and 0.3 ± 0.3 [77]. Discretisation gains 7.5 points at L1 and 18.5 at L2 against within-condition spreads of ±1.9 and ±0.6, so the ranking survives a spread test at those two levels; at L3 a 3.4-point signal against roughly ±1.5 is marginal, and at L4 nothing is measured. It points opposite to the Chinese working consensus, which prices 256 bins and a 50-step horizon against 3 to 5 Hz autoregressive inference and 0.33 s per timestep on an A100 [41].

Chunk horizon is open. Every located source specifies a horizon and none sweeps it against success; this review located no horizon-against-performance curve, having searched the Chinese, Japanese, Korean, German and French literatures in native language and native venues. One Japanese diffusion policy states the horizon fully, observation 2, prediction 16, executed 5, at 50 Hz control [95]. The Japanese answer elsewhere is that the question is malformed: a two-rate hierarchy runs 2.5 Hz upper against 50 Hz lower over a 500 Hz control loop [89], and the null arm predicts the next step only, at 25 Hz under zero-order hold while the 500 Hz controller runs [87]. Halving the control period from 2 ms to 1 ms bought up to 30% faster task execution [91]. SmolVLA prices the execution schedule rather than the horizon: asynchronous inference holds success near 78% while cutting completion from 13.75 s to 9.7 s, and doubles tasks completed in fixed time, 19 against 9 [85].

The emitter family has one like-for-like comparison on identical real hardware, and it is confounded. On RoboChallenge's Table30, task-specific finetunes reach SR 43.7% for pi0.5, 28.3% for pi0 and 11.7% for CogACT under one grading rule [17], so flow-matching emitters beat the diffusion-action VLA by 3.73x and 2.42x. Those systems differ in backbone, pretraining corpus and scale, so the comparison ranks models rather than action representations. The only common-interface comparison located must not be ranked: MLP 16.4 ± 15, ACT 45.0 ± 18, Diffusion Policy 52.0 ± 30 and SARNN 50.8 ± 25 over 8 MuJoCo tasks and 5 seeds, where a 7.0-point between-method signal sits against ±18 to ±30 within-method spread [98]. Only MLP separates. Energy descent holds one real-hardware number, 71% zero-shot grasp success trained exclusively on simple simulated objects, with no matched diffusion or flow baseline [62].

The compute term is why the binding constraint on action representation is computation efficiency. Against a 20 ms budget at 50 Hz, one configuration of eight clears the deadline on mean and maximum, streaming DDIM/16 with TensorRT at 8.43 and 10.23 ms, against 352.38 and 375.15 ms for DDPM/100 without it [96]; the inference-engineering span of 41.8x on means and 36.7x on maxima [97] exceeds any accuracy difference between emitter families here. The transition-function gap sits at roughly 30 degrees of ninety, binding constraint AI algorithms, and moves when a published model reaches dynamic consistency near 0.9 at scene parity, against a benchmark maximum of 0.6197 at scene 0.9436 [24]. The action-representation gap sits at roughly 38 degrees of ninety, and moves when an emitter clears 20 ms worst case without streaming, or when the control target moves to the 1 kHz a commodity research interface already exposes [54], cutting the budget to 1 ms.

Quantity	Value	Grade	What it constrains
EWMBench dynamics at scene parity, OpenSora against EnerVerse_FT	Scene 0.9210 against 0.9427; dynamic consistency 0.0474 against 0.5363, an 11.3x gap	verified [24]	Whether a world model scored on pixels says anything about action conditioning
RoboChallenge "temporal" tag, 3 tasks, 10 rollouts per task	SR 5%, progress score 14, against an all-task mean of SR 22% and score 37	verified [16]	The measured cost of single-frame policies carrying no transition function
Physical-consistency loss on the transition, real robot	82.2% to 94.4%, on 15 and 39 demonstrations, 10 rollouts per distance, untrained 24 cm included	verified [90]	That a prior on the transition buys real success at small demonstration counts
Stiffness emitted inside the predicted t+20 ms next state	15 against 10 Nm/rad discriminated from 5 demonstrations per bottle	verified [92]	Whether an estimated material parameter can condition the transition online
Discretised classification against continuous	L1 90.8 against 83.3; L2 47.8 against 29.3; within-condition	verified [77]	The head choice, in simulation

regression, identical backbone	spreads ±1.9 and ±0.6		only; no hardware extension located
Chinese working consensus on discretisation	256 bins, 50-step horizon, 3 to 5 Hz autoregressive against 50 Hz required, 0.33 s per timestep on an A100	reported [41]	The opposite conclusion to the row above, stated without an in-document protocol
Emitter families on identical real hardware, Table30	pi0.5 SR 43.7%, pi0 28.3%, CogACT 11.7%	verified [17]	Family ranking, confounded by backbone, pretraining corpus and scale
Common-interface policy comparison, 8 MuJoCo tasks, 5 seeds	MLP 16.4 ± 15, ACT 45.0 ± 18, Diffusion Policy 52.0 ± 30, SARNN 50.8 ± 25	self-published [98]	Nothing beyond separating MLP; within-method spread exceeds between-method signal
Energy-descent policy on real hardware	71% zero-shot grasp success, trained exclusively on simple simulated objects	verified [62]	The single real-hardware number the third emitter family holds, with no matched baseline
Diffusion inference against a 20 ms budget at 50 Hz	1 of 8 configurations clears mean and maximum; streaming DDIM/16 with TensorRT 8.43 / 10.23 ms	reported [96]	The deadline that decides the emitter before accuracy does
Inference-engineering span across those eight configurations	41.8x on means, 36.7x on maxima	modelled [97]	That engineering, rather than family, dominates the choice in this dataset
Chunk horizon as specified, never swept	Observation 2, prediction 16,	reported [95]	The one fully stated horizon located; no

	executed 5, at 50 Hz control		horizon-against-success curve was located
Certified Cartesian speed against effective robot mass, chest	1500 down to 400 mm/s from 1 to 20 kg, a 3.75x contraction	verified [50]	That the certified action set is state-dependent, which no located representation encodes
Korean national target for world-model contribution to real success	At or above 20 pp, against a stated global best of 14.5 pp	self-published [12]	Nothing yet: no protocol, N or task list is attached to either number
Run-to-run variance with task, props and model held constant	0% to 100%	verified [15]	Every ranking above

Quantities bearing on the action-conditioned transition function and on action representation, with grade and the design decision each one constrains.

6. The sim-to-real tax as a stateable quantity, and drift on the envelope against drift on the task

Four bands of appearance randomisation, priced in real success on a UR5 over seven tasks by twenty episodes: two-dimensional image augmentation alone scores 0/20 on every task; adding 1,203 ambientCG textures on robot, table, wall and floor gives 11.3/20, or 56.5%; adding lighting and object colour gives 14.0/20, or 70.0%; adding camera parameters gives 18.6/20, or 93.0% [71]. The bands are published with the ladder, including the 1,203 texture assets and a light-source distance between 1.0 and 3.0 m from the workspace centre [72]. Nothing in the ladder randomises mass, friction, latency or any actuator term. Against 98.34% for the same policy family over 250 simulation episodes per task, the tax is 5.34 percentage points [74]. That figure carries its own caution. Twenty real trials per task resolves to 5 pp per task, the same size as the tax being reported.

The Korean band sweep runs three bands at 10,000 episodes each and reports simulation-converged grasp success: DR1 at 97% or better within about 1,000

steps, DR2 at about 94% stable after about 3,000 steps, DR3 at 86 to 89% after 10,000 steps, with early-training variance of about 0.18 [2]. The real endpoint on a UR16e is a single figure, 77.3% rising to 91.7% over 65 trials [1]. Differencing the two sides gives 8.7 to 11.7 pp against the hardest band and 19.7 pp against the easiest [3], which is the conservation shape: wider bands, smaller drop, paid for in simulated performance. This review reads the pair as refuted on the quantity. Both endpoints are real-world figures on the same robot, policy pool and simulator, differing only in the ranking rule, and the three tiers vary object spacing and obstacle count rather than randomisation width [2].

One German design runs a single evaluation protocol in simulation and on a KUKA LBR iiwa over 50 identical target poses, with the band-setting rule stated: latency in 0 to 1 s, joint stiffness and damping each in [1,65], observation noise in 0 to 10%, each band widened until an agent trained in an ideal simulator degraded by more than 10% [59]. Best-model mean reward without randomisation runs -72.03 in simulation to -142.50 on hardware, and with full randomisation -96.01 to -109.10 [60]. Modelled from those four numbers, the reality gap is 97.8% without randomisation and 13.6% with it, and the price paid in simulation is 33.3% [61]. The trade then inverts with operating point. At pi/9 rad/s the same design gives -79.05 to -129.10 without randomisation and -158.69 to -197.22 with it [60]. Full randomisation loses 101% in simulation and lands worse on hardware than no randomisation at all [61].

The widest bands in the record ship with no performance accounting this review located. One vendor configuration sets friction_range to 0.1 to 1.25 against a base default of 0.5 to 1.25, added_mass_range to -1.0 to +3.0 kg against base -1 to +1, push_interval_s to 5 against base 15, with observation noise scales of 0.01 on joint position and 1.5 on joint velocity [14]. Where five axes are swept together rather than laddered, the paired real number is 9.0% rising to 42.0% zero-shot in cluttered scenes with unseen backgrounds [26]. A cable-driven soft manipulator with a 648 mm backbone transfers zero-shot at 43.5 mm mean absolute error, 6.71% of length, against millimetre-order accuracy inside simulation [79]. Gap 3 stands at roughly 40 degrees of ninety, and the binding constraint is engineering implementation. The threshold quantity is a third joint-speed point in the German design; while -197.22 and -129.10 straddle -109.10 and -142.50, no conservation statement can be made.

Drift has been a certified quantity at exactly one corner of the envelope since 1998. ISO 9283:1998 carries drift of pose accuracy and drift of pose repeatability, and in its Tables 1 and 2 the drift test is the only characteristic specified solely at 100% rated load and 100% rated velocity, with no optional second operating condition,

where every other characteristic has one [64]. The rest of the envelope side is dense and equally silent on tasks. One research arm's datasheet states a safe Cartesian position accuracy worst case at stop of 50 mm against a point repeatability better than plus or minus 0.1 mm, a 500:1 ratio in one document on one arm [53]. Reglement (UE) 2023/1230 sets permitted drift to zero, learning phase included, and subjects self-evolving safety components to third-party conformity assessment from 20 January 2027 [83]. Japan defers force, torque and pressure thresholds to individual risk assessment [108]. Korea admits analysis-only certification as one of four accepted verification routes [8].

Measurement on the task has one retrievable instance carrying a full protocol. Holding task, robot and policy fixed, moving one nuisance variable, and running 20 trials per cell, the randomisation-trained policy scores 20/20 by default, 20/20 on a textured tablecloth, 19/20 in low light, 16/20 under multicolour light, 18/20 on object colour and 11/20 under camera variation, mean 17.3/20; the real-data-trained policy scores 20/20, 1/20, 17/20, 14/20, 19/20 and 8/20, mean 13.2/20 [75]. The textured-tablecloth column, 100% against 5%, is the cleanest measurement of drift on the task located in either sweep. One national standard mandates scene-disturbance decay rate as a required per-task metric [21]; this review did not locate a numeric acceptance value or a published per-company score behind it, having searched the issuing ministry's release, the state news relay and CAICT-sourced trade press. Gap 7 stands at roughly 35 degrees of ninety, binding constraint regulatory.

Gap 6, a reference implementation of a drift protocol, stands at roughly 15 degrees of ninety, and the binding constraint is engineering implementation. Five regional sweeps did not locate one, having searched in Chinese, Japanese, Korean, German and French against native venues. What blocks it is a variance term rather than a missing instrument. With the same props, the same task and the same model held fixed, a measured real success rate can move from 0% to 100% or the reverse, depending on which class of human resets the scene [15]. The published fix superimposes a reference frame from a held-out episode on the live camera preview, and this review did not locate a residual spread under it. A published residual of 10 percentage points or better at that benchmark's own 10 rollouts per task [17] would make a drift protocol implementable on an already-published control list [43]. Certification on the envelope is not measurement on the task.

randomisation band	measured real success	region	grade

2D image augmentation only	0/20 on every one of 7 tasks, UR5, 20 episodes per task	French	verified [71]
plus 1,203 ambientCG textures on robot, table, wall, floor	11.3/20, 56.5%	French	verified [71]
plus lighting and object colour	14.0/20, 70.0%	French	verified [71]
plus camera parameters	18.6/20, 93.0%	French	verified [71]
none (ideal simulator), 1 rad/s joint speed	mean reward −142.50, KUKA LBR iiwa, 50 identical target poses	German	verified [60]
latency 0 to 1 s, stiffness and damping, observation noise 0 to 10%, 1 rad/s	mean reward −109.10	German	verified [1,65 60]
none (ideal simulator), pi/9 rad/s	mean reward −129.10	German	verified [60]
same three bands, pi/9 rad/s	mean reward −197.22	German	verified [60]
no extensive randomisation, simple simulated objects only	71% real grasp success	German	verified [62]
none, zero-shot SOFA/Cosserat transfer	43.5 mm mean absolute position error, 6.71% of a 648 mm backbone	French	verified [79]
five axes: clutter, lighting, background, tabletop height, language	9.0% rising to 42.0% zero-shot, cluttered real scenes, unseen backgrounds	Chinese	verified [26]
shipped vendor config: friction_range 0.1 to 1.25, added_mass_range −1.0 to +3.0 kg, push_interval_s 5	this review did not locate a real success figure accompanying these bands, having searched the vendor source tree	Chinese	verified [14]

DR1, DR2, DR3 tiers (object spacing and obstacle count, not band width)	77.3% on a UR16e over 65 trials, one real endpoint for all three tiers	Korean	verified [1]
same tiers, simulation minus real	8.7 to 11.7 pp against the hardest band, 19.7 pp against the easiest	Korean	modelled [3]

Randomisation band against measured real success, five regions. Reward-denominated rows are not commensurable with success-rate rows, and only the German study runs one evaluation protocol on both sides.

7. Contact, friction and deformables

ISO/TS 15066:2016 Annex A is a published pressure and energy field with its protocol attached. Table A.2 states maximum permissible peak pressure for 29 localisations, from 110 N/cm² at the temple to 300 N/cm² at the dominant index-finger pad [46], and quasi-static force for twelve regions, from 65 N at the face to 220 N at thigh and knee, with a transient multiplier of 2 everywhere except head and face, where transient contact is not permitted [47]. Table A.4 states maximum transferable energy per region, 0.11 J at the face to 2.6 J at the pelvis [49]. The protocol travels with the numbers: 100 subjects, pain-onset threshold, third quartile, a 1.4 by 1.4 cm metal stamp with a 2 mm edge radius, pressure averaged at 1 mm² resolution [46]. This review located no policy-learning paper in any of the eleven regions treating the 0.11 J face limit as a constraint on a learned contact policy, having searched all five in the native language.

Gap 11 sits at 78 degrees of ninety, and the binding constraint is regulatory. The phantom is buildable today: DGUV FB HM-080 Tabelle 1 specifies damping material, thickness and spring constant per body region, with the neck at 70 Shore A over 7 mm against 50 N/mm, and the same document reports a measured null, that exchanging the springs moves the force results little [52]. What moves the position is the field itself. The forehead pair stands at 130 N and 130 N/cm² [46,47], 6.5x the area pressure of the prior German national recommendation of 90 N at 20 N/cm², with no new physiology in between [66]. Regions disagree on one finger: Korean guidance sets the index finger at 520 N/cm² and 500 N [7] against 300 N/cm² and a 140 N hand-and-finger force in the German retrieval [46,47]. What closes is the human-biomechanical field alone; this review did not locate a holding whose subject is a differentiable contact model in any of the five sweeps.

The fingertip half is priced and specified in four regions, and its resolution premise fails. Minsight measures 0.07 N mean absolute force error and 0.6 mm contact-location error over 1,740 mm² at 60 Hz; Insight before it gives 0.03 N and 0.4 mm at 11 Hz, and across nine sensors tabulated in the same paper nothing reaches better than 0.4 mm [57]. FingerVision runs at 30 to 60 fps at a material cost near USD 50 [105]; XELA's uSPa 21 resolves about 0.98 mN per axis at 500 Hz [104]; Leptrino's six-axis line runs from JPY 107,800 to JPY 1,232,000 at 1.2 kHz [103]. Chinese vendors quote DM-Tac W at 40,000 sensing units per cm² for ¥1,299, and PaXini from ¥199 at 0.01 N and 1 kHz [33], an 18.1x spread inside one functional class [34]. Korean vendors add a 6 mN tactile sensor [9] and a hand skin detecting 0.01 N at about KRW 30,000 per sensor [10].

The whole-body half is open, and gap 4 sits at 40 degrees of ninety with engineering implementation binding. TUM's H-1 carries 1,260 hexagonal cells and more than 13,000 sensors, with event-driven processing reduction stated as up to 90%, and this review located no price, no per-taxel force resolution and no calibration protocol for it [58]. Chinese whole-body coverage runs to 16-channel and 64-channel arrays, 1,000-plus channels at best [35], against 40,000 units per cm² at the fingertip [33], the two halves counted in different units, channels per body region against units per cm², and this review located no interface between them in any of the five sweeps. Certification pulls the other way, requiring at least 1 kHz [51] against tacterion's 120 Hz taroki [56]. A third route bypasses the instrument, since CRANE-X7 recovers disturbance torque by observer at a 500 Hz control period with no torque sensor in the arm [88]. The threshold is one vendor document in any region stating taxel count, per-taxel force resolution, unit price and calibration protocol together.

Friction is estimated in three regions and always on the wrong side of the contact. Korean work measures viscous friction coefficients of quadruped leg joints online, 0.052 and 0.092 in air against 0.112 and 0.156 in water, with knee-pitch tracking error integral falling from 0.0585 to 0.0197 [4,5]. Japanese work identifies forward Coulomb 0.11966 Nm and forward viscous 2.10020 Ns by simulated annealing at 10 kHz sampling [94]. The Nantes school runs two-stage IDIM-LS over exciting trajectories at 250 Hz [84]. All three yield Nm and Ns per joint, never a dimensionless μ. Exactly one surface μ was located in eleven regions, a swept ratio Ft/FN taken during instrument characterisation with acquisition capped at 2.5 Hz and no identified coefficient [81]. The assumption is made in the open: Unitree's shipped G1

configuration carries a uniform friction prior of 0.1 to 1.25 [14], and German gripper practice publishes $\mu = 0.1$ with a safety factor of 2 [63].

Sensing is not the binding constraint on gap 5. Chinese laboratories report 6 mN normal-force accuracy, 0.21 mm mean shape error on a curved palm and 98.98% slip detection [36]; slip is detected as a binary event and never inverted into a coefficient. A Japanese rig measures plantar friction force directly, 830 N maximum normal at the heel and 320 N maximum friction at the big toe for a 90 kg male, reporting newtons rather than a ratio, with the two peaks at different anatomical sites [93]. The standards carry no friction term: ISO/TS 15066 Annex A is entirely pressure and force [46,47], and the Japanese limit set carries 250 mm/s and the former 150 N guideline [107]. The field's leading real-robot benchmark measures no force at all [18]. Gap 5 sits at 30 degrees of ninety, binding constraint engineering implementation, and the threshold is that 0.1 to 1.25 prior replaced by an estimated μ with a stated variance.

Four non-finite-element architectures each carry a retrievable number for gap 12. An implicit neural representation reaches 75% on untangling a twisted rubber band against a prior best of 26% [13]. A Cosserat model in SOFA transfers zero-shot to a 648 mm cable-driven prototype at 43.5 mm mean absolute error, 6.71% of length [79]. An online scalar stiffness inside a predicted t+20 ms next state discriminates 15 from 10 Nm/rad on two PET bottles from 5 demonstrations each [92]. The fourth is the 1-DOF spring certification accepts, 10 to 150 N/mm [48]. Capability stays open on the field's own protocol: RoboChallenge's softbody tag sits at 8% success and progress score 27 against a 30-task mean of 22% [16], while the strongest task-specific policy there reaches 43.7% [17]. Gap 12 sits at 33 degrees of ninety, binding constraint AI algorithms; the threshold is 90% coverage on 20 unseen deformable classes within 300 s, over 20 repetitions [42].

Body region or sensor	Quantity	Value	Grade
Face (ISO/TS 15066 Table A.4)	Maximum transferable energy	0.11 J	verified [49]
Pelvis (ISO/TS 15066 Table A.4)	Maximum transferable energy	2.6 J	verified [49]
Face (ISO/TS 15066 Table A.2)	Quasi-static force limit; transient	65 N	verified [47]

	contact not permitted		
Index-finger pad (ISO/TS 15066 Table A.2)	Peak pressure, dominant / non-dominant	300 / 270 N/cm²	verified [46]
Index finger (Korean collaborative-work guide)	Allowable pressure and force, quasi-static	520 N/cm², 500 N	verified [7]
Abdomen and skull (ISO/TS 15066 Table A.3)	Effective spring constant, low and high ends	10 N/mm and 150 N/mm	verified [48]
Neck test body (DGUV FB HM-080 Tabelle 1)	Durometer, thickness, spring constant	70 Shore A, 7 mm, 50 N/mm	verified [52]
Contact measurement chain (DGUV FB HM-080)	Minimum sampling frequency and filter	1 kHz, 100 Hz Butterworth	verified [51]
Minsight fingertip (Germany)	Force error / contact-location error at 60 Hz	0.07 N / 0.6 mm over 1,740 mm²	verified [57]
Insight fingertip (Germany)	Force error / contact-location error at 11 Hz	0.03 N / 0.4 mm over 4,800 mm²	verified [57]
FingerVision (Japan)	Frame rate and material cost	30 to 60 fps, about USD 50	verified [105]
XELA uSPa 21 (Japan)	Force resolution per axis and sampling rate	0.98 mN at 500 Hz	self-published [104]
Leptrino six-axis line (Japan)	List price at 1.2 kHz sampling	JPY 107,800 to 1,232,000	self-published [103]
PaXini PX-6AX-GEN3 (China)	Force resolution, output rate, entry price	0.01 N, 1 kHz, ¥199	self-published [33]
DM-Tac W (China)			

	Sensing density, output rate, list price	40,000 units per cm², 120 Hz, ¥1,299	self-published [33]
Fingertip tactile class, China	Price spread inside one functional class	18.1x	modelled [34]
Aidin ATT tactile sensor (Korea)	Force resolution at 5.5 mm thickness	6 mN	self-published [9]
Holiday Robotics hand skin (Korea)	Detectable force and per-sensor cost	0.01 N at about KRW 30,000	self-published [10]
TUM H-1 whole-body skin (Germany)	Cell and sensor count	1,260 cells, more than 13,000 sensors	reported [58]
tacterion taroki (Germany)	Tactile sampling rate	120 Hz	self-published [56]
Chinese electronic skin arrays	Channels per body region	16 to 64, 1,000-plus at best	reported [35]
Human fingertip (French reference set)	Minimum detectable pressure and spatial resolution	about 1 mN, 0.5 mm	reported [82]
Quadruped leg joints (Korea)	Viscous friction coefficient, air then water	0.052 / 0.092 then 0.112 / 0.156	verified [4]
Unitree G1 shipped configuration (China)	Ground-friction prior as a uniform band	0.1 to 1.25	verified [14]
German gripper design practice	Design friction coefficient and safety factor	μ = 0.1, factor 2	reported [63]
Heel and big toe (Japanese plantar rig)	Maximum normal / maximum friction force, 90 kg male	830 N / 320 N	verified [93]
		8%, score 27	

RoboChallenge softbody tag (China)	Success rate and progress score, 10 rollouts per task		verified [16]
Cable-driven soft manipulator (France)	Zero-shot mean absolute position error	43.5 mm, 6.71% of length	verified [79]
PET bottles, JRSJ 44(4) (Japan)	Stiffness discriminated from 5 demonstrations each	15 versus 10 Nm/rad	verified [92]

Contact limits, tactile instruments and friction quantities located across the five regional sweeps, each figure with the grade it carries.

8. Teleoperation data, the real-to-sim loop, and system identification

Hours per usable teleoperated demonstration is held at verified grade in none of the eleven regions swept, and the eleven yield four incompatible denominators. The one candidate was discarded on verification: 150 real demonstrations in approximately 4.5 hours on a UR5, or 108 s per demonstration [73], proved to be autonomous scripted-oracle replay of a simulator's action list, the word teleoperation does not occur in that chapter, and 4.5 hours carries one significant figure. What survives at verified grade is demonstration counts with their protocols, and all of it is Japanese: 48 demonstrations over 6 initial layouts by 8 trials for a six-step contact-rich assembly [87]; 27 human demonstrations bootstrapped sixfold to 162 by folding successful autonomous rollouts back in [89]; and 15 and 39 demonstrations on two tasks, against a stated prior requirement of 404 data points for comparable motion generation [90]. Position on gap 8: 40 degrees of ninety. Binding constraint: workforce.

China has an occupational category for the work. 数采 is a named industry segment with the job titles 采集员 and 具身数据师, operating out of 训练场, warehouse-scale facilities running shifts against throughput targets, at least fifteen of them, against national demand estimated at 100,000 to 200,000 hours per year [28]. The day rate is ¥300 per collector-day for 2 to 5 hours of usable demonstration data and roughly 500 trajectories at a pass rate above 90%, on rigs above ¥200,000, against a market price of ¥500 to ¥1,000 per hour [28]. Labour alone is therefore ¥60 to ¥150 per hour of data and roughly ¥0.67 per usable trajectory [29]. Against roughly 500,000 hours held industry-wide and roughly 10,000,000 stated as required, the 20x shortfall

costs ¥0.6 to 1.5bn in collector wages alone [30]. No definition of usable and no yield protocol accompany any of it.

The threshold on gap 8 is that same collector-day figure at verified grade, with a stated definition of usable and a yield protocol, because every modelled price above moves linearly with it. Korea states the metric pair as a national target rather than a measurement: one hour of continuous collection at 80% success or better, with the binding constraint named as operator labour intensity [11]. Two Chinese results pull in opposite directions on the same question. FastUMI removes the robot body, and with it the force channel, cutting single-trajectory collection from about 50 s to about 10 s [31]. ForceMimic keeps force in the demonstration and measures it as worth 54.5% over vision-only imitation on contact-rich tasks, applying a mean 7.5 N against the human demonstrator's 6 N [32]. This review located no published crossover point between the two routes.

Gap 9 is closed at one-degree-of-freedom actuator scale by exactly one retrievable document, and the Japanese identification cost sweep is what closes it. Shimoichi and Katsura match a simulator's position response to the 実機 by simulated annealing and publish the full cost curve against cooling rate: 0.9 gives 2,989 trials at objective 0.022276, 0.99 gives 28,762 trials at 0.004156, and 0.999 gives 279,927 trials at 0.000451 [94]. Ninety-four times the compute buys forty-nine times the accuracy. The identified set at the finest setting is forward Coulomb friction 0.11966 Nm, backward 0.07018 Nm, forward viscous 2.10020 Ns, backward 0.58302 Ns and inertia 0.00745 kg·m², sampled at 10 kHz, with residual error concentrated at 0.2 s and 1.2 s and attributed to backlash absent from the model [94]. It covers one joint, five parameters, no contact, and it is fourteen years old.

At task scale no located source states a loop fidelity with its protocol. Shanghai AI Laboratory reports navigation rising from 50% zero-shot to 80% with real-to-sim data, and manipulation from 46.7% to 93.3% at a 5:1 simulation-to-real 数据配比, at a forward-looking ¥0.02 per synthesised trajectory against roughly ¥10 per real one [27]. ISW Stuttgart runs a closed loop for limp objects with both rates stated, DART physics at 1 ms against a stereo camera at 100 Hz, and this review did not locate an accuracy figure for it [67]. ABB claims simulation accuracy of about 99 percent from a virtual controller running identical firmware, with no task, no protocol and no distribution [69]. France prices the loop instead: 39 real samples plus simulation hold 8.7 mm, 27 break to 10.1 mm, and 39 real samples alone give 18.3 mm [80]. Position on gap 9: 30 degrees of ninety. Binding constraint: engineering

implementation. Threshold: a published loop holding accuracy below 27 real samples.

System identification priced in energy was searched for in eleven regions, in native terms against native venues, and this review did not locate a single source in any of them that prices it in joules. The lever is priced instead in five mutually incommensurable currencies. Korea prices it in control bandwidth and bytes: a 20 kHz torque loop, a 2 kHz angular-velocity loop and a 12-byte FD-CAN uplink at 4 Mbps, carrying recursive least squares at forgetting factor 0.99 plus a Gopinath observer inside the driver [6]. China prices it in seconds and percent: a five-term Fourier excitation trajectory at 0.04 Hz fundamental over 20 periods, 60 s total at 100 Hz sampling, torque prediction error within 5% [23]. Japan prices it in optimiser trials [94]. France states 250 Hz sampling along excitation trajectories, and this review did not locate a cost figure attached to it [84].

The instrument to price identification in joules sits in Germany, and this review did not locate it pointed at an identification run: 1.0 to 3.0 kWh average consumption by payload class, controller standby of 30 to 140 Wh, and a worked example of 9.95 Wh per cycle at 14.64 s and 2446 W [68]. Every one of those attaches to a production cycle. Joules are already the currency of contact safety, from 0.11 J at the face to 2.6 J at the pelvis [49]. Cross-region arithmetic gives the shape of the missing experiment rather than a result: 80 W of controller draw [54] across 60 s of excitation [23] is 4.8 kJ, a floor only, since it excludes the motor work the German record shows dominating by roughly thirty times. The denominator sits unused: 3 kWh over more than 6 h implies roughly 500 W mean draw [45]. Position on gap 10: 8 degrees of ninety. Binding constraint: regulatory. Threshold: whether 平均任务能耗 in YD/T 6770-2026 charges calibration and identification to the task budget [21].

Quantity	Value	Region	Grade
Usable demonstration data per collector-day	2 to 5 hours, roughly 500 trajectories, pass rate above 90%	Chinese	reported [28]
Collector day rate against data market price	¥300 per person-day against ¥500 to ¥1,000 per hour	Chinese	reported [28]
Labour cost per hour and per usable trajectory	¥60 to ¥150 per hour; roughly ¥0.67 per usable trajectory	Chinese	modelled [29]

Industry data shortfall and its wage bill	500,000 h held against 10,000,000 h required; ¥0.6 to 1.5bn in wages	Chinese	modelled [30]
Single-trajectory collection time, embodiment-free rig	about 50 s to about 10 s, cost cut to one fifth	Chinese	self-published [31]
Measured worth of force in the demonstration	+54.5% over vision-only; 7.5 N applied against the human's 6 N	Chinese	reported [32]
National teleoperation collection target	1 hour continuous at 80% success or better	Korean	self-published [11]
Demonstration budget for a six-step contact task	48 demonstrations, 6 layouts by 8 trials	Japanese	verified [87]
Teleoperation throughput correction	43% faster than no correction, 50% over the prior method, about 88% on the hardest target	Japanese	verified [110]
Identification cost sweep against annealing cooling rate	2,989 / 28,762 / 279,927 trials at objective 0.022276 / 0.004156 / 0.000451	Japanese	verified [94]
Classical excitation-trajectory identification cost	60 s at 100 Hz sampling, torque prediction error within 5%	Chinese	reported [23]
Online identification cost, stated in bandwidth	20 kHz torque loop, 2 kHz velocity loop, 12 bytes at 4 Mbps	Korean	verified [6]
Real samples at which the real-to-sim loop stops paying	N=39 gives 8.7 mm, N=27 gives 10.1 mm, 39 real alone gives 18.3 mm	French	verified [80]
Simulation-to-real mixing ratio and its success pair	5:1 ratio, 46.7% to 93.3%; ¥10 to ¥0.02 per trajectory	Chinese	self-published [27]

Closed real-to-sim loop rates for limp objects	DART physics at 1 ms against stereo vision at 100 Hz, a 10:1 ratio	German	reported [67]
Robot work priced in joules, production cycle	9.95 Wh per cycle at 14.64 s and 2446 W	German	reported [68]
Energy floor for one classical excitation run	80 W controller draw across 60 s of excitation = 4.8 kJ, a floor only	German and Chinese	modelled on reported inputs [54] [23]

Quantities bearing on gaps 8, 9 and 10: teleoperation data, the real-to-sim loop, and system identification. Bracket numbers index the review's reference list.

9. Transfer evidence and who evaluates it

RoboChallenge measured a real-robot success rate moving from 0% to 100%, with the same props, the same task and the same model held fixed; the only thing that changed was which class of human reset the scene between rollouts [15]. Three classes ran two tasks, stacking bowls and pouring fries: experienced testers who had collected the training data, ignorant testers on first exposure, and adaptive testers who were the model's own authors. Adaptive testers scored highest, and the report documents a sweet-spot effect in which the adaptive tester located and reused favourable box positions [15]. That is a variance term wider than almost every effect reported anywhere in this material. It applies to the rankings below, to the rankings this review drew in earlier sections, and to any future Institute number. A ranking is admissible only where the between-item signal exceeds the within-cell spread. This review did not locate a published residual spread under RoboChallenge's own fix, Visual Task Reproduction, having searched the arXiv paper and its relays [15].

Third-party evaluation, meaning an evaluator who is not the model's author, was located in one region of five. RoboChallenge scores models it did not build: pi0.5 task-specific finetune at 43.7% success and 62.2 progress score, pi0 at 28.3%, CogACT at 11.7%, and pi0.5 under the generalist protocol at 17.7%, a factor of 0.41 against its own task-specific number [17]. By difficulty tag, temporal tasks sit at 5% and softbody at 8% [16]. The annual report covers more than 40,000 real-robot trials across 30 standardised tasks, with top-three average success in the 35 to 51% band and one later entry at 64.33% [40]. RoboDojo runs 18 real tasks at 10 trials each, 180

trials per policy, each scored double-blind by three reviewers, with human teleoperators at 100.0% against a best policy of 12.8% [19]. Two reproducible control lists are published, with named control variables [43]. CAICT's second evaluation batch is running; this review did not locate published per-company scores, having searched the programme's own platform [44].

Everywhere else the evaluator is the claimant, so self-publication is the wrong discriminator; trial count and protocol completeness are the right ones. AIST scores its own three baselines on 8 real UR5e tasks at 43.8% ±38, 47.9% ±32 and 39.6% ±38 [99]; the between-method signal is 8.3 points against spreads of ±32 to ±38, so the three policies are not separated by that experiment [101], each cell rests on 6 rollouts [100], and the underlying real dataset is listed as coming soon [102]. A Japanese peer-reviewed letter reports per-step success across 6 layouts at 5 rollouts per cell [86]. A French thesis reports 8.1 out of 10 against 6.7 out of 10 on real hardware, where one trial is one of the ten points [78]. One located study ran a single protocol on both sides, 50 identical target poses in simulation and on a KUKA LBR iiwa [60]. A J-STAGE query for Sim-to-Real with 実機 returns 4,862 results, so any claim of no external evidence would be a search failure [112].

China's most-quoted transfer pair is 89.4% success in simulation against 12% in real household scenes, a stated 77-point transfer chasm, attributed across at least six Chinese outlets to the Stanford HAI AI Index Report 2026 [20]. This review did not locate those two figures in that report, having searched the AI Index document itself and each Chinese relay; no relay cites a page or a chapter [20]. The pair is also inconsistent with the paired real-hardware evidence. RoboDojo's real leaderboard tops out at 12.8%, and its 42 simulation tasks and 18 real tasks are different task sets, so 8.80% against 12.8% is not a transfer measurement [19]; AIST's simulation and real task lists are likewise disjoint [101]. No source located in eleven regions publishes a paired sim-and-real success measurement on one task set. Any use of 77 points in this corpus is unsourced twice over.

Two further numbers get read as transfer claims and are not. Eight humanoids ran 64 cumulative hours on a tablet quality-control line, 17,625 units, at a stated 99.99% operation success under continuous live stream [37]; the arithmetic reconciles to about 7 failed operations in 55,000 to 65,000 repetitions of a fixed motion on a fixed fixture, a failure rate roughly 7,800 times apart from RoboDojo's 12.8% per-task figure [38]. The difference is the task envelope, not the policy, and any Institute reliability claim must state which of the two it is. Separately, a vendor reports 98% zero-shot first-grasp success with no real-machine training data across 100 or more

unseen objects; the reports located state no trial count and no reset procedure, and this review did not locate a protocol [39]. That is the precise configuration under which a measured success rate was shown to move from 0% to 100% [15].

This review struck its own strongest Korean candidate. The 77.3% to 91.7% grasp pair over 65 real trials [1] is not a transfer measurement, because both endpoints are real-world figures differing only in the ranking rule, and the stated trial count does not produce either percentage as a denominator. Nothing located evaluates the Institute's own policies, and the only SO-100-class transfer number in the five-region record is 51.7% to 78.3%, with the trial count not stated [85]. Gap 13 sits at 30 degrees of ninety. The binding constraint is engineering implementation: the control lists exist [43] and the hardware runs are what is missing, one national coordination budget outside China being EUR 20 M over four years across fourteen institutions [70], roughly 600 arms at the EUR 33,180 list price [55]. Threshold: 30 rollouts per policy-task cell on SO-101-class hardware under a published control list, with within-cell spread below the claimed effect.

Claim	Evaluator	Independent of the claimant	Grade
Real-robot success rate moves 0% to 100% with task, props and model fixed	RoboChallenge (Dexmal, Hugging Face)	Yes; the instrument measured on itself	verified [15]
pi0.5 43.7%, pi0 28.3%, CogACT 11.7% success on 30 real tasks	RoboChallenge	Yes	verified [17]
Temporal tag 5%, softbody tag 8%, all-task mean 22%	RoboChallenge	Yes	verified [16]
More than 40,000 real trials; top-three average 35 to 51%	RoboChallenge annual report, via trade relay	Yes	reported [40]
Human teleoperators 100.0% against best policy 12.8%, 18 real tasks	RoboDojo, three reviewers double-blind	Yes	reported [19]
Paired sim and real reward over 50		No	

identical target poses, KUKA LBR iiwa	Josifovski et al., on their own policies		verified [60]
Real success 0/20 rising to 18.6/20 as randomisation bands are added, UR5	Inria/Willow thesis author, own method	No	verified [71]
ACT 43.8%, Diffusion Policy 47.9%, SARNN 39.6% on 8 real UR5e tasks	AIST, on its own baselines	No	self-published [99]
6 rollouts per real cell, inferred from percentage granularity	This review's arithmetic on Table IV	No	modelled [100]
89.4% simulation against 12% real household, 77-point gap	At least six outlets, attributed to AI Index 2026	Not located in the attributed document	reported [20]
99.99% operation success, 64 h, 17,625 units, live-streamed	Zhiyuan (智元) and Longcheer (⍰旗科技)	No	self-published [37]
98% zero-shot first-grasp success, no real-machine training data	Sudo (⍰度科技)	No	self-published [39]
51.7% to 78.3% on SO100 with pretraining, trial count not stated	LeRobot team, on its own model	No	self-published [85]
77.3% to 91.7% grasp success over 65 real trials	Paper's own authors; struck by this review	No	verified [1]

Transfer claims located in the five-region sweep, with the evaluator, whether that evaluator is independent of the party making the claim, and the grade carried in this review's reference index.

10. Position and binding constraint, gap by gap, across eleven regions

Thirteen gaps in the Institute's corpus were stated before either sweep ran. Version 1 adjudicated them against five regions: one closed, two open, ten partly. Re-adjudicated against eleven, every one of the thirteen moved in position, in binding constraint, or in both. That is the strongest argument in this review for reading more of the world's record before concluding anything about it.

Three verdicts change in kind rather than degree. Gap 11 moves from closed to partly closed, for the reasons in Section 2. Gap 1 changes from no such holding exists to four such holdings exist, each modelling one machine on one surface, with no shared evaluation protocol, so they cannot be composed. Gap 12's literal claim does not survive at all: four physics families now appear with real hardware in at least three regions, and the binding constraint moves from AI algorithms to comparability, since none of the four shares a benchmark with another.

The two gaps that stayed open both sharpened. A reference implementation of a drift protocol is still not located, and the constraint moves from no implementation exists to an in-force national instrument exists and contains no threshold. System identification priced in energy is still not located as a measured quantity, and the constraint moves from regulatory to instrumentation: the one region that measured robot joules at all was limited by its meter. Both are now better-specified openings than they were, which is what a sweep is for.

The binding constraints themselves moved more than the positions did. Across the thirteen, seven changed class. Several moved away from technology and toward measurement: gap 3 from engineering implementation to measurement protocol, gap 4 to measurement discipline, gap 8 from workforce to the absence of a shared denominator, gap 13 from self-publication to pairing and statistical power. One moved to incentive, gap 5, because the field has a working substitute in slip detection plus randomisation and nobody is forced to measure the coefficient. And gap 7's constraint became commercial confidentiality rather than regulatory absence, which is the first time in either sweep that a constraint landed there.

Gap	v1, five regions	v2, eleven regions	Binding constraint now
	partly	four holdings exist, uncomposable	AI algorithms

1 Transition function as subject			
2 Action representation	partly	a fifth option appears	actuation-side rate and smoothness
3 The sim-to-real tax	partly	seven numbers agreeing in neither sign, magnitude nor unit	measurement protocol
4 The tactile epidermis	partly	force half is a commodity in four markets	measurement discipline
5 Friction estimated, not assumed	partly	still not located in eleven	incentive: a working substitute exists
6 A reference drift implementation	open	open, sharpened	an in-force instrument with no threshold
7 Envelope versus task	partly	moved in three places outside the literature	commercial confidentiality
8 Hours per usable demonstration	partly	four incompatible denominators	absence of a shared denominator
9 A closed real-to-sim loop	partly	four loops in three regions	the closure criterion
10 Identification priced in energy	open	open, sharpened	instrumentation: the meter
11 Human-biomechanical contact field	closed	partly closed	regulatory, with the container moved
12 Deformables	partly	literal claim does not survive	comparability
13 Third-party evaluation of transfer	partly	58 of 146 sweep-two findings bear on it	pairing and statistical power

Table 10. All thirteen gaps at five regions and at eleven. Every row moved. Positions are the authors' assessment against the cited quantities and are not measurements.

11. What this review did not locate, and what that is worth

Across both sweeps, 74 quantities in the first and a further register in the second were searched for and not located. Eleven of them were searched for in every one of the eleven regions and located in none: a measured in-situ contact coefficient of friction between an end effector and a manipulated object; system identification priced in joules, which is located in seconds, core-hours, parameters, currency and product options and in joules nowhere; a numeric adequacy threshold for a simulation standing in for a physical test; an event-driven whole-body skin, every whole-body instrument located being polled between 1 and 500 Hz; a tactile or force datasheet publishing measured accuracy beside quoted resolution, other than one device in one region; a shared denominator for teleoperation data; a sim-to-real tax that keeps its sign under a change of success predicate, object set or method; a confidence interval on a real-world transfer trial count stated by the authors rather than computed here; a national humanoid programme with staged numeric task-success targets other than one; and a published sensor price in four of the currencies swept.

How much stronger is eleven than five. Less than it looks, and in one respect considerably more. Less, because the eleven are not eleven independent literatures. Every one of the six regions in the second sweep reports, in its own words, that its frontier work is published in English at the venue set this corpus already reads. An absence found in all eleven is therefore substantially one literature observed eleven times rather than eleven independent draws, and reporting it as an elevenfold replication would overstate it by an unknown but large factor. This review does not make that claim.

More, because the second sweep replaced failed searches with measured censuses in four places. A hand count of 2,159 titles across four Brazilian proceedings. Roughly 270 Spanish papers searched in full text with a soundness cross-check on the index. Exact-phrase counts against a named Russian full-text index on a stated date. Three Indian journals enumerated by source identifier with the recent output listed item by item. A census is a different epistemic object from a search that returned

nothing, and it is why the Brazilian and Spanish nulls are usable where an ordinary null would not be.

What these statements are not. They are statements about what two sweeps located, in named indexes, on named dates, using named query vocabularies. They are not statements about what exists. Access failures are recorded as access failures and not as absences: a per-body-region reproduction of the contact table in a language other than German sits behind a paywall in most regions and is filed accordingly. And eleven regions is not the world. Arabic, Persian, Turkish, Indonesian, Vietnamese, Thai, Polish, Czech, Ukrainian, Dutch, Greek and Hebrew were not swept, and nothing here should be read as a statement about those records.

Quantity not located	Regions that searched and did not find it	Standing
In-situ contact coefficient of friction, measured	11 of 11	Gap 5 stays open in practice
System identification priced in joules	11 of 11	Gap 10 stays open; constraint is the meter
Numeric adequacy threshold for simulation standing in for test	11 of 11	Gap 6 stays open; an instrument exists with no threshold
An event-driven whole-body skin	11 of 11	Every instrument located is polled
Measured accuracy beside quoted resolution on a datasheet	10 of 11	One device in one region publishes both
A shared denominator for teleoperation data	11 of 11	Four incompatible units
A sim-to-real tax that keeps its sign	11 of 11	The tax is predicate-dependent, not a constant
Contact table reproduced per body region outside German	most regions	Filed as ACCESS, not absence: paywalled

Table 11. What neither sweep located. The final row is an access failure and is not evidence about the world, which is the distinction this review exists to hold.

References

1. [1] 임수빈·이재선, 「시뮬레이션-실환경 통합 점수 기반 로봇 정책 선택 기법」(Unified Sim-and-Real Scoring Methods for Robot Policy Selection), 로봇학회논문지 (The Journal of Korea Robotics Society) Vol. 20, No. 3, pp.371-380, print 29 Aug 2025, DOI 10.7746/jkros.2025.20.3.371. SKKU 대학원 + KITECH 안산. Retrieved full text from jkros.org. https://www.jkros.org/_PR/view/?aidx=46105&bidx=4378
   VERIFIED
2. [2] 임수빈·이재선, 로봇학회논문지 20(3):371-380, 2025. §4.3 「시뮬레이터 실험」, [Table 3] Domain Configuration. https://www.jkros.org/_PR/view/?aidx=46105&bidx=4378
   VERIFIED
3. [3] My arithmetic on the sim (§4.3) and real (§4.4) figures in 임수빈·이재선, 로봇학회논문지 20(3):371-380, 2025. The paper itself does not compute this difference. https://www.jkros.org/_PR/view/?aidx=46105&bidx=4378
   MODELLED
4. [4] 박재형·안재훈·류호균·강호선·송화영·김민성·이인호, 「모터 동역학 기반 실시간 점성 마찰 추정 및 보상을 통한 보행 로봇의 적응 제어」(Real-Time Viscous Friction Estimation and Compensation Based on Motor Dynamics for Adaptive Control of Legged Robots), 로봇학회논문지 20(3):495-503, 2025, DOI 10.7746/jkros.2025.20.3.495. 부산대학교. [Table 3]. Retrieved PDF. https://jkros.org/xml/46120/46120.pdf
   VERIFIED
5. [5] 박재형 외, 로봇학회논문지 20(3):495-503, 2025. §4.1: "특히, TEI가 0.0585에서 0.0197로 2배 이상 개선된 것을 볼 수 있다. 반면, HP에서는 TEI 값이 약간 높아졌으나…". https://jkros.org/xml/46120/46120.pdf
   VERIFIED
6. [6] 박재형 외, 로봇학회논문지 20(3):495-503, 2025. §2.1.1, [Table 1] Robot System Parameter, [Table 2] Control and Estimator Parameter. Funded by IITP RS-2023-00215760 (안내 로봇 Guide Dog). https://jkros.org/xml/46120/46120.pdf
   VERIFIED
7. [7] 「산업용 로봇의 협동작업 안전 가이드」(Industrial Robot Collaborative-Work Safety Guide), 2023. 7., Korean government guidance superseding the 2022.9 edition; 부록 1 협동작업 형태별 기능사항, 동력 및 힘 제한 (PFL) section, p.27. The table's cited authority is the Korean standard KOROS 1162-1 「인간-로봇 접촉에 대한 생체 역학적 임계치」. Retrieved and text-extracted from the distributed PDF. https://www.smartcona.co.kr/upload/board/img_file_board_file_1725422116.pdf
   VERIFIED
8. [8] 「산업용 로봇의 협동작업 안전 가이드」, 2023. 7., 부록 1, 동력 및 힘 제한 items 1 and 4. https://www.smartcona.co.kr/upload/board/img_file_board_file_1725422116.pdf

VERIFIED

9. [9] 「[기업 최전선을 가다-에이딘로보틱스] 로봇에 '촉각'을 입히다」, 로봇신문 (irobotnews), article no. 43143. Figures originate with 에이딘로보틱스. https://www.irobotnews.com/news/articleView.html?idxno=43143
SELF-PUBLISHED

10. [10] 「"3만원 센서로 만든 휴머노이드 손…내년 양산 개시"」, ZDNet Korea, 2025-11-03. Figures originate with 홀리데이로보틱스. https://zdnet.co.kr/view/?no=20251103104411
SELF-PUBLISHED

11. [11] 「[AI클로즈업] K-휴머노이드 착수… 실시간 데이터 연결·모델 경량화 관건」, 디지털데일리, 발행 2026-05-21, 구아현 기자. Targets originate with 박재흥 서울대 교수팀 under the MSIT/KIST 민관협력 기반 AI 휴머노이드 원천기술 고도화 사업. Raw Korean retrieved and verified. https://www.ddaily.co.kr/page/view/2026052016512856045
SELF-PUBLISHED

12. [12] 「[테크 해설] 유리 반사와 너머를 갈라냈다…340억 월드모델의 목표 20%p」, 이포커스, 입력 2026.07.12, 곽도훈 기자, citing 과학기술정보통신부·IITP (2026.6.9.). Raw Korean retrieved and verified. https://www.e-focus.co.kr/news/articleView.html?idxno=3002546
SELF-PUBLISHED

13. [13] KAIST 연구뉴스, INR-DOM. 제1저자 송민석 (석사과정), 교신 박대형 교수 (전산학부). Presented at Robotics: Science and Systems (RSS) 2025, June 21-25, USC Los Angeles. https://www.kaist.ac.kr/researchnews/html/news/?mode=V&mng_no=50671&skey=keyword&sval=INR-DOM
REPORTED

14. [14] unitree_rl_gym, legged_gym/envs/g1/g1_config.py and legged_gym/envs/base/legged_robot_config.py (宇树科技官方强化学习仓库), retrieved from GitHub raw, 2026-08-12. https://raw.githubusercontent.com/unitreerobotics/unitree_rl_gym/main/legged_gym/envs/g1/g1_config.py
VERIFIED

15. [15] RoboChallenge: Large-scale Real-robot Evaluation of Embodied Policies (Dexmal 北京原力灵机 + Hugging Face), arXiv:2510.17950, 2025-10-20; PDF retrieved and read directly. https://arxiv.org/pdf/2510.17950
VERIFIED

16. [16] RoboChallenge: Large-scale Real-robot Evaluation of Embodied Policies, Table 2, arXiv:2510.17950, 2025-10-20. https://arxiv.org/pdf/2510.17950
VERIFIED

17. [17] RoboChallenge: Large-scale Real-robot Evaluation of Embodied Policies, Figure 9 / Sec 3.2, arXiv:2510.17950, 2025-10-20. https://arxiv.org/pdf/2510.17950
VERIFIED

18. [18] RoboChallenge: Large-scale Real-robot Evaluation of Embodied Policies, Sec 2.2, arXiv:2510.17950, 2025-10-20. https://arxiv.org/pdf/2510.17950
VERIFIED

19. [19] RoboDojo: A Unified Sim-and-Real Benchmark for Comprehensive Evaluation of Generalist Robot Manipulation Policies (港大MMLab / 清华 / 北大 与合作机构; RoboTwin 原班团队), arXiv:2607.04434, v1 2026-07-05, v3 2026-07-08; 中文报道题为「具身智能『高考』难疯了！人类100分，最强模型12.8」, 量子位, 2026-07-08. https://arxiv.org/abs/2607.04434
REPORTED

20. [20] 「77%的『迁移鸿沟』如何跨越？真机数据成为具身智能行业新赛点」, 机器人科技网 / robotsci.com.cn, 2026; and 「人形机器人交付元年，行业从卷模型转向拼数据」, 麻省理工科技评论中国版 (MIT Technology Review China), 2026-05. http://www.robotsci.com.cn/detail/2052227568328904704
REPORTED

21. [21] 「具身智能领域首份行业标准发布——推动技术成果产业化应用」, 中国政府网 gov.cn, 2026-03-30; 「加速产业规模化 具身智能标准持续完善」, 新华网, 2026-06-01; 「6月1日起，具身智能领域新标准正式实施」, 通信世界网, 2026. https://www.gov.cn/lianbo/202603/content_7064102.htm
VERIFIED

22. [22] 国家标准|GB/T 40575-2021, 国家标准全文公开系统 openstd.samr.gov.cn, retrieved 2026-08-12. https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=5DF8297C5C12E67D5F870D2F5F001AAD
VERIFIED

23. [23] 「一种带负载工业机器人动力学模型辨识方法」, 南京航空航天大学学报 (Journal of Nanjing University of Aeronautics and Astronautics), 2016 年第6期; and 「采用ABC算法的关节机器人动力学参数辨识」, 同刊 2017 年第5期. https://jnuaa.nuaa.edu.cn/html/2016/6/201606009.htm
REPORTED

24. [24] EWMBench: Evaluating Scene, Motion, and Semantic Quality in Embodied World Models (智元机器人 AgiBot: Hu Yue, Siyuan Huang, Yue Liao, Shengcong Chen, Pengfei Zhou, Liliang Chen, Maoqing Yao, Guanghui Ren), arXiv:2505.09694, 2025-05; HTML tables read directly. https://arxiv.org/html/2505.09694v1
VERIFIED

25. [25] 「智元主办AGIBOT WORLD CHALLENGE 2026收官 比赛结果公布」, 网易科技 relaying 智元机器人 organiser release, 2026; 「AGIBOT WORLD CHALLENGE冠军：世界模型需从画面逼真转向动作可控」, 新浪财经, 2026. https://www.163.com/dy/article/KUU69HR705118A6A.html
SELF-PUBLISHED

26. [26] RoboTwin 2.0: A Scalable Data Generator and Benchmark with Strong Domain Randomization for Robust Bimanual Robotic Manipulation (上海交大 / 港大 / 上海人工智能实验室; Tianxing Chen ... Xiaokang Yang, Ping Luo, Yao Mu), arXiv:2506.18088, 2025-06-22; abstract retrieved from arXiv API. https://arxiv.org/abs/2506.18088
VERIFIED

27. [27] 「上海AI实验室构建具身智能『虚实贯通』技术体系，数字伙伴DigitalBuddy助力降本增效」, 上海人工智能实验室 (Shanghai AI Laboratory) 官方新闻, 2025-01-26. https://www.shlab.org.cn/news/5444053
SELF-PUBLISHED

28. [28] 《2026中国具身智能数据采集与数据产业发展展望》, 亿欧智库 (EqualOcean Intelligence), 2026; 「具身数据采集产业链调查：被机器人采集的人」, 投资界/36氪, 2026-06; 「具身智能带火了数据采集生意」, 21世纪经济报道, 2026-05-07. https://news.pedaily.cn/202606/565591.shtml
REPORTED

29. [29] my arithmetic on 亿欧智库《2026中国具身智能数据采集与数据产业发展展望》(¥300/day, ¥500-1000/h) and 投资界/36氪 industry survey (2-5 h/collector-day, >90% pass, ~500 trajectories/day), 2026-08-12. https://news.pedaily.cn/202606/565591.shtml
MODELLED

30. [30] my arithmetic on 亿欧智库《2026中国具身智能数据采集与数据产业发展展望》(500k h existing / 10M h required; ¥300/day; ¥500-1000/h) and 投资界/36氪 (2-5 h/collector-day), 2026-08-12. https://www.fxbaogao.com/detail/5506283
MODELLED

31. [31] 「具身智能的数据难题，终于有了可规模化的解法」, 量子位, 2025-12; and 「具身智能数据战开打！每个普通人都能上手，边采边筛，只投喂机器人爱吃的丨穹彻」, 量子位, 2026-01. https://www.qbitai.com/2025/12/361743.html
SELF-PUBLISHED

32. [32] ForceMimic: Force-Centric Imitation Learning with Force-Motion Capture System for Contact-Rich Manipulation (上海交通大学卢策吾团队 + 穹彻智能 Noematrix), arXiv:2410.07554, 2024-10; Chinese release 「复杂连续接触操作：穹彻智能与上交大共创规模化力反馈模仿数据与学习模型-力捕捉系统ForceMimic」, 2024-12-11. https://www.alphaxiv.org/zh/overview/2410.07554v1
REPORTED

33. [33] 「全球首款！戴盟家族产品正式发布，加速具身智能产业化进程」, 量子位, 2025-04 (DM-Tac W); 「极致性价比，性能再升级！PaXini 第三代多维触觉传感器矩阵全球耀世首发」, 机器人大讲堂, 2026 and 帕西尼商城 mall.paxini.com (GEN3, ¥199 起); 「灵巧手制作成本太高，突破口在哪？」, 观察者网风闻, 2026 (¥3,600 ITPU fingertip). https://leaderobot.com/news/6015
SELF-PUBLISHED

34. [34] my arithmetic on 帕西尼 GEN3 pricing (机器人大讲堂 / mall.paxini.com, 2026) and the ¥3,600 ITPU fingertip figure (观察者网风闻, 2026). https://leaderobot.com/news/6015
MODELLED

35. [35] 「人形机器人触觉传感器，爆发前夜」, Supplyframe 四方维 / 新浪财经, 2026-02-26; 「2026电子皮肤行业报告：具身智能感知核心」, 2026; 「宇树机器人深度解析：Dex3-1灵巧手深度解析！」, 艾邦机器人. https://finance.sina.com.cn/roll/2026-02-26/doc-inhpcqki0499646.shtml

REPORTED

36. [36] 「机器手力度识别精度达 6 mN，中国科大在触觉传感器研究中取得重要进展」, 新浪科技 relaying 中国科学技术大学新闻网, 2024-11-26; 「国内首篇！融合语言模型的多模态触觉传感器，推动机器人触觉迈向人类水平」, 新浪科技 relaying 清华大学, 2026-01-25; 「新颖的曲面手掌视触觉传感器」, 浙江大学学报(工学版), 2025; 「基于注意力机制和视触融合的机器人抓取滑动检测」, 信息与控制. https://news.ustc.edu.cn/info/1048/89951.htm
REPORTED

37. [37] 「打破两个全球纪录，智元第1.5万台机器人下线，6天直播成功率99.99%」, 新浪财经, 2026-06-29; 「连续6天进厂，8台智元机器人直播『打工』，再启工业级能力验证」, 投资界, 2026-06. https://news.pedaily.cn/202606/565641.shtml
SELF-PUBLISHED

38. [38] my arithmetic on 智元/龙旗 day-5 and final figures (新浪财经 2026-06-29; 投资界 2026-06) and RoboDojo real-world leaderboard (arXiv:2607.04434). https://arxiv.org/abs/2607.04434
MODELLED

39. [39] 「20亿刀苏度科技具身首秀：0真机数据，0 shot，首次抓取成功率98%」, 新浪财经, 2026-04-20; 「『仿真派』落地真产线！苏度WAIC首秀，CEO韩铮：99%+成功率」, 网易/智源社区, 2026. https://finance.sina.com.cn/stock/t/2026-04-20/doc-inhvcmef8644397.shtml
SELF-PUBLISHED

40. [40] 「一份来自40000+次真机评测的具身智能年度报告！RoboChallenge打破Demo滤镜：最强模型也只有51%成功率」, CSDN资讯, 2026-02; 「基于数万次真机评测，机器人大模型年度评测报告发布」, 中国日报网, 2026-02-04; 「新纪录！星动纪元Era0登顶全球具身智能真机评测榜」, 光明网, 2026-05-20. https://cn.chinadaily.com.cn/a/202602/04/WS6982f0c6a310942cc499e3b1.html
REPORTED

41. [41] 「《VLA 系列》OpenVLA-OFT | 并行解码 | 连续动作 | VLA动作解码方法对比」, CSDN, 2026; 「The Compression Gap：为什么discrete tokenization阻碍 VLA 模型性能提升？」, 知乎, 2026; 「Xbotics论文日报（2026.03.30）| DFM-VLA：从『一次生成』到『逐步精化』的离散动作解码框架」, 知乎, 2026-03-30. https://zhuanlan.zhihu.com/p/2028576766915036981
REPORTED

42. [42] 「ICRA 2026 | 李飞飞团队：软物体移动操作新解法，『从刚到柔』的关键一步」, 网易, 2026 (protocol); 「世界第一！让全体AI翻车的叠衣难题，被这家中国实验室拿下」, 智源社区/新浪财经, 2026-06 (LeHome Challenge); 「叠衣见真章：深圳具身智能从『炫技』迈向『实用』新阶段」, 中国金融信息网, 2026-06-01. https://hub.baai.ac.cn/view/55337
REPORTED

43. [43] RoboDojo (arXiv:2607.04434) and 「RoboTwin原班团队再下场，构建具身评测珠峰」, 网易/智源社区, 2026-07; RoboChallenge Sec 2.3.2 and 2.3.4, arXiv:2510.17950. https://arxiv.org/abs/2607.04434
REPORTED

44. [44] 「中国信通院具身智能基准测试（EAI Bench）第二批评测（Y2026Q1）工作正式启动！」, 鲸智社区·大模型公共服务平台 (CAICT), 2026; 「中国信通院牵头推进具身智能国际标准工作推动两项新立提案」, 新浪财经/第一财经, 2025-11-17; EIBench 百度百科条目. https://aihub.caict.ac.cn/docs/7E6SoNH5ZZvh
REPORTED

45. [45] 「最多续航4小时 人形机器人『电量焦虑』何时解？」, 证券时报网 stcn.com / 新浪财经, 2026-07-31; 「续航仅120分钟的人形机器人，技术瓶颈突破的时间窗还剩多久？」, 新浪, 2026-06-05. https://www.stcn.com/article/detail/4050928.html
REPORTED

46. [46] ISO/TS 15066:2016(E), Robots and robotic devices — Collaborative robots, Annex A, Table A.2 „Biomechanical limits"; deutsche Wiedergabe als Anhang A in DGUV-Information FB HM-080 „Kollaborierende Robotersysteme – Planung von Anlagen mit der Funktion Leistungs- und Kraftbegrenzung", Ausgabe 08/2017. Beide Volltexte abgerufen. https://www.dguv.de/medien/fb-holzundmetall/publikationen-dokumente/infoblaetter/infobl_deutsch/080_roboter.pdf
VERIFIED

47. [47] ISO/TS 15066:2016(E), Table A.2; DGUV-Information FB HM-080, Anhang A, Ausgabe 08/2017, Anmerkung 2. https://www.dguv.de/medien/fb-holzundmetall/publikationen-dokumente/infoblaetter/infobl_deutsch/080_roboter.pdf
VERIFIED

48. [48] ISO/TS 15066:2016(E), Table A.3 „Effective masses and spring constants for the body model". https://www.diag.uniroma1.it/deluca/pHRI_elective/ISO_TS_15066_2016_en.pdf
VERIFIED

49. [49] ISO/TS 15066:2016(E), Table A.4 „Energy limit values based on the body region model". https://www.diag.uniroma1.it/deluca/pHRI_elective/ISO_TS_15066_2016_en.pdf
VERIFIED

50. [50] ISO/TS 15066:2016(E), Table A.5 „Example of calculated transient contact speed limit values based on the body model". https://www.diag.uniroma1.it/deluca/pHRI_elective/ISO_TS_15066_2016_en.pdf
VERIFIED

51. [51] DGUV-Information FB HM-080 „Kollaborierende Robotersysteme – Planung von Anlagen mit der Funktion ‚Leistungs- und Kraftbegrenzung'", Fachbereich Holz und Metall der DGUV, Ausgabe 08/2017, Seite 5. https://www.dguv.de/medien/fb-holzundmetall/publikationen-dokumente/infoblaetter/infobl_deutsch/080_roboter.pdf
VERIFIED

52. [52] DGUV-Information FB HM-080, Ausgabe 08/2017, Tabelle 1 „Dämpfungsmaterial und Federkonstanten für Messanordnung nach Bild 7". https://www.dguv.de/medien/fb-holzundmetall/publikationen-dokumente/infoblaetter/infobl_deutsch/080_roboter.pdf
VERIFIED

53. [53] „Datenblatt FRANKA RESEARCH 3", Franka Robotics GmbH, Version 2.1, Juni 2025, Dokumentennummer R02212 (Herstellerdatenblatt, PDF abgerufen). https://franka.de/hubfs/Digital_Datasheet%20Franka%20Research%203_R02212_2.1_DE.pdf
REPORTED

54. [54] „Datenblatt FRANKA RESEARCH 3", Franka Robotics GmbH, Version 2.1, Juni 2025, R02212. https://franka.de/hubfs/Digital_Datasheet%20Franka%20Research%203_R02212_2.1_DE.pdf
REPORTED

55. [55] Generation Robots, Produktseite „7-axis Franka Research 3 Robotic Arm (FCI licence)"; Unchained Robotics, Produktseite „Bota Systems SensONE Cobot Kit - EtherCAT" (beide abgerufen 12.08.2026). https://unchainedrobotics.de/en/products/end-of-arm-effectors/end-of-arm-accessories/force-torque-sensors/bota-systems-sensone-cobot-kit-ethercat
REPORTED

56. [56] taroki® – Taktiles Sensor-Kit für industrielle Roboter, tacterion GmbH, München (Herstellerseite, deutsch). https://www.taroki.de/german
SELF-PUBLISHED

57. [57] Andrussow, Sun, Kuchenbecker, Martius, „Minsight: A Fingertip-Sized Vision-Based Tactile Sensor for Robotic Manipulation", Advanced Intelligent Systems 5(8), 2023, doi 10.1002/aisy.202300042; Max-Planck-Institut für Intelligente Systeme, Stuttgart (arXiv:2304.10990 abgerufen). https://arxiv.org/pdf/2304.10990
VERIFIED

58. [58] „Sensible Roboter sind sicherer: Verbesserte Wahrnehmung durch biologisch inspirierte künstliche Haut", Pressemitteilung TU München / MIBE, verweisend auf G. Cheng et al., Proceedings of the IEEE, 2019, doi 10.1109/JPROC.2019.2933348 (Lehrstuhl für Kognitive Systeme, TUM). https://www.tum.de/en/news-and-events/all-news/press-releases/details/35732-1
REPORTED

59. [59] Josifovski, Malmir, Klarmann, Žagar, Navarro-Guerrero, Knoll, „Analysis of Randomization Effects on Sim2Real Transfer in Reinforcement Learning for Robotic Manipulation Tasks", IEEE/RSJ IROS 2022, Kyoto (TU München; TH Rosenheim; L3S/DFKI). arXiv:2206.06282 abgerufen. https://arxiv.org/pdf/2206.06282
VERIFIED

60. [60] Josifovski et al., IROS 2022, Tabellen III und IV (arXiv:2206.06282). https://arxiv.org/pdf/2206.06282
VERIFIED

61. [61] Eigene Arithmetik auf Josifovski et al., IROS 2022, Tabellen III und IV. https://arxiv.org/pdf/2206.06282
MODELLED

62. [62] Gergely Sóti, „Robotic Grasping via Implicit Policies on Neural Radiance Fields", Dissertation, Institut für Anthropomatik und Robotik (IAR), Karlsruher Institut für

Technologie, 2026 (KITopen 1000191293). https://publikationen.bibliothek.kit.edu/1000191293
VERIFIED

63. [63] „Greifkraftberechnung", IPR – Intelligente Peripherien für Roboter GmbH, Schwaigern (Herstellerseite, deutsch); ergänzend Festo Info 139 „Parallelgreifer HGPT, HGPL und Dreipunktgreifer". https://www.iprworldwide.com/greifkraftberechnung/
REPORTED

64. [64] ISO 9283:1998, „Manipulating industrial robots — Performance criteria and related test methods", Tabellen 1 und 2 (Vorschau-PDF von standards.iteh.ai abgerufen); Prüfdauer von 8 h aus Sekundärquellen zu Klausel 7.6, nicht im abgerufenen Auszug enthalten. https://cdn.standards.iteh.ai/samples/22244/d058a9d8bacf41cd85def67ab70afc87/ISO-9283-1998.pdf
VERIFIED

65. [65] ISO/PAS 5672:2023, „Robotics — Collaborative applications — Test methods for measuring forces and pressures in human-robot contacts", ISO/TC 299, Abschnitt 5.3 (Vorschau-PDF von standards.iteh.ai abgerufen); deutschsprachige Einordnung über die Wissensdatenbank rokit des Fraunhofer IFF. https://cdn.standards.iteh.ai/samples/82488/e9f2f066763b4189b18e2b9d3e6dccec/ISO-PAS-5672-2023.pdf
VERIFIED

66. [66] IFA-Fachinfos „Kollaborierende Roboter (COBOTS) – Medizinisch/Biomechanische Anforderungen" und Projektbeschreibung BGIA 5105, Deutsche Gesetzliche Unfallversicherung (DGUV); Zahlen aus Sekundärzusammenfassung, nicht aus dem Volltext U 001/2009 abgerufen. https://www.dguv.de/ifa/forschung/projektverzeichnis/bgia_5105.jsp
REPORTED

67. [67] „Verformte Teile? Kein Problem – Roboterdemonstrator kombiniert Simulation und Bildverarbeitung", ROBOTIK UND PRODUKTION, 29.05.2020; Institut für Steuerungstechnik der Werkzeugmaschinen und Fertigungseinrichtungen (ISW), Universität Stuttgart, DFG-Graduiertenkolleg Soft Tissue Robotics. https://robotik-produktion.de/allgemein/verformte-teile-kein-problem/
REPORTED

68. [68] „Mit Robotern energieeffizient produzieren", Beschaffung aktuell, 02.09.2013; „Energiebewertungsmethode", Fraunhofer IWU Chemnitz (Projektseite, deutsch); „Fraunhofer IWU: GreenBotAI senkt Energieverbrauch von Robotern um 25 Prozent", heise online, 10.04.2024. https://www.iwu.fraunhofer.de/de/projekte/roboter-energiebewertungsmethode.html
REPORTED

69. [69] „Physical AI in der Industrie: Wie ABB Robotics und NVIDIA die Lücke zwischen Simulation und Realität schließen", ABB Destination Zukunft (deutschsprachige Herstellerpublikation, 2026). https://destination-zukunft.abb.com/robotik/physical-

ai-in-der-industrie-abb-robotics-und-nvidia-schliessen-luecke-zwischen-simulation-und-realitaet/
SELF-PUBLISHED

70. [70] „Robotics Institute Germany (RIG) treibt KI-basierte Robotik in Deutschland voran", KIT-Pressemitteilung 06/2024; BMFTR-Bekanntmachung „Robotics Institute Germany (RIG)", 23.11.2023. https://www.kit.edu/kit/202406-robotics-institute-germany-treibt-ki-basierte-robotik-in-deutschland-voran.php
REPORTED

71. [71] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation » (thèse, Centre Inria de Paris / Willow, résumé bilingue), HAL tel-05535510, 2025 — Table 3.3; also published as Garcia et al., IROS 2023. https://hal.science/tel-05535510v1
VERIFIED

72. [72] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation », HAL tel-05535510, 2025, §3 — identical to Garcia, Strudel, Chen, Arlaud, Laptev, Schmid, « Robust Visual Sim-to-Real Transfer for Robotic Manipulation », arXiv:2307.15320. https://arxiv.org/abs/2307.15320
VERIFIED

73. [73] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation », HAL tel-05535510, 2025, §3 (« gathering 150 demonstrations on the real robot is time-consuming (approx. 4.5 hours) »). https://hal.science/tel-05535510v1
VERIFIED

74. [74] my arithmetic on Tables 3.1 and 3.3 of Ricardo Garcia-Pinel, HAL tel-05535510, 2025. https://hal.science/tel-05535510v1
MODELLED

75. [75] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation », HAL tel-05535510, 2025, Table 3.4. https://hal.science/tel-05535510v1
VERIFIED

76. [76] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation », HAL tel-05535510, 2025, Table 5.4; benchmark published ICRA 2025. https://hal.science/tel-05535510v1
VERIFIED

77. [77] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation », HAL tel-05535510, 2025, Table 5.5. https://hal.science/tel-05535510v1
VERIFIED

78. [78] Ricardo Garcia-Pinel, « Learning Visuomotor Policies for Robotic Manipulation », HAL tel-05535510, 2025, §5.5 and §4. https://hal.science/tel-05535510v1
VERIFIED

79. [79] Guillaume Samain, « Conception et commande d'un manipulateur souple actionné par câbles : approche par apprentissage et transfert de la simulation à la réalité » (thèse, Institut des Systèmes Intelligents et de Robotique, Sorbonne), HAL tel-05526023, 2025, Tableau 5.1. https://theses.hal.science/tel-05526023v1

VERIFIED

80. [80] Guillaume Samain, HAL tel-05526023, 2025, Figures 5.4–5.7, Tableaux 5.1–5.2. https://theses.hal.science/tel-05526023v1
VERIFIED

81. [81] Lionel Fliegans, « Études sur les transducteurs multimodaux optiques souples pour la perception robotique et la restauration des fonctions sensorielles tactiles » (thèse, École Nationale Supérieure des Mines de Saint-Étienne, NNT 2024EMSEM030), HAL tel-05019306, 2024. https://theses.hal.science/tel-05019306v1
VERIFIED

82. [82] Lionel Fliegans, thèse Mines Saint-Étienne, HAL tel-05019306, 2024, chapitre 2 (citing refs [10],[12],[20],[26],[27],[28]). https://theses.hal.science/tel-05019306v1
REPORTED

83. [83] Règlement (UE) 2023/1230 (Règlement Machines), cité verbatim dans UNM/AFNOR, « Guide pratique sur la notion d'IA telle qu'utilisée dans le Règlement Machines 2023/1230 », 2025; et Pierre Belingard (EUROGIP), « Les nouveaux risques machines à travers le nouveau règlement RM 2023/1230 », journée technique INRS, 25 mars 2025. https://eurogip.fr/contributions/guide-pratique-sur-la-notion-dia-telle-quutilisee-dans-le-reglement-machines-2023-1230/
VERIFIED

84. [84] Pauline Hamon, Maxime Gautier, « Identification dynamique de robots avec un modèle de frottement sec fonction de la charge et de la vitesse », CIFA/conférence, Nancy, juin 2010, HAL hal-00583159; et Claire Dumas, « Identification et simulation physique d'un robot Stäubli TX90 pour le fraisage à grande vitesse », thèse HAL tel-00831071. https://hal.science/hal-00583159v1
REPORTED

85. [85] Hugging Face (équipe LeRobot, Paris), « SmolVLA: Efficient Vision-Language-Action Model trained on LeRobot Community Data », blog + arXiv 2506.01844, 2025. https://huggingface.co/blog/smolvla
SELF-PUBLISHED

86. [86] 山根広暉, 境野翔, 辻俊明「バイラテラル制御に基づく模倣学習による複数物体の同時把持」日本ロボット学会誌 Vol.42 No.4, pp.394-397, 2024年5月 (レター, 査読付, 有用性(システム設計・構築分野)評価). https://www.jstage.jst.go.jp/article/jrsj/42/4/42_42_394/_pdf/-char/ja
VERIFIED

87. [87] 同上 日本ロボット学会誌 42(4) 2024, §4.2 データセット / §4.3 学習モデル. https://www.jstage.jst.go.jp/article/jrsj/42/4/42_42_394/_pdf/-char/ja
VERIFIED

88. [88] 日本ロボット学会誌 42(4) 2024 §3.2, and 41(4) 2023 §3.2 (same group, same platform). https://www.jstage.jst.go.jp/article/jrsj/42/4/42_42_394/_pdf/-char/ja
VERIFIED

89. [89] 日本ロボット学会誌 41(4) 2023 §4.2-4.3, 5.結言. https://www.jstage.jst.go.jp/article/jrsj/41/4/41_41_395/_pdf/-char/ja
VERIFIED

90. [90] 佐藤寛, 桝屋望, 山根広暉, 稲見洸紀, 佐藤涼穂, 境野翔, 辻俊明「角度と角速度の関係性を考慮したバイラテラル制御に基づく模倣学習」日本ロボット学会誌 Vol.43 No.6, pp.603-606, 2025. https://www.jstage.jst.go.jp/article/jrsj/43/6/43_43_603/_pdf/-char/ja
VERIFIED

91. [91] 稲見洸紀, 大明準治, 境野翔, 辻俊明「制御周期の短縮によるバイラテラル制御の高精度化と学習応用の可能性」日本ロボット学会誌 Vol.44 No.3, p.332, 2026. https://www.jstage.jst.go.jp/article/jrsj/44/3/44_44_332/_article/-char/ja
VERIFIED

92. [92] 「物体の剛性情報を利用した柔軟物把持の模倣学習」日本ロボット学会誌 Vol.44 No.4, pp.433-436, 2026年5月. https://www.jstage.jst.go.jp/article/jrsj/44/4/44_44_433/_pdf
VERIFIED

93. [93] 「足裏の垂直抗力と摩擦力を測定・提示可能なシステムの開発」日本ロボット学会誌 Vol.44 No.5, pp.512-515, 2026年6月. https://www.jstage.jst.go.jp/article/jrsj/44/5/44_44_512/_pdf
VERIFIED

94. [94] 下市拓真, 桂誠一郎「実世界ハプティクスのための最適化問題に基づく摩擦同定法」計測自動制御学会論文集 Vol.48 No.12, pp.907-912, 2012年12月, DOI 10.9746/sicetr.48.907. https://www.jstage.jst.go.jp/article/sicetr/48/12/48_907/_pdf/-char/ja
VERIFIED

95. [95] 同上 JSAI2026 4I1-GS-8b-05, §3-4. https://www.jstage.jst.go.jp/article/pjsai/JSAI2026/0/JSAI2026_4I1GS8b05/_pdf/-char/ja
REPORTED

96. [96] 同上 JSAI2026 4I1-GS-8b-05, 表1 推論時間の比較. https://www.jstage.jst.go.jp/article/pjsai/JSAI2026/0/JSAI2026_4I1GS8b05/_pdf/-char/ja
REPORTED

97. [97] my arithmetic on 表1, JSAI2026 4I1-GS-8b-05. https://www.jstage.jst.go.jp/article/pjsai/JSAI2026/0/JSAI2026_4I1GS8b05/_pdf/-char/ja
MODELLED

98. [98] 産業技術総合研究所 (AIST) 「RoboManipBaselines: A Software Framework for Imitation Learning in Robotic Manipulation across Real and Simulation Environments」 arXiv:2509.17057, TABLE II. https://arxiv.org/pdf/2509.17057
SELF-PUBLISHED

99. [99] AIST, arXiv:2509.17057, TABLE IV. https://arxiv.org/pdf/2509.17057
SELF-PUBLISHED

100. [100] my arithmetic on TABLE IV granularity, cross-checked against the repo's stated sim protocol ("rollouts at six different object positions for each of the five

checkpoints trained with different random seeds"). https://github.com/isri-aist/RoboManipBaselines/blob/master/doc/evaluation_results.md
MODELLED

101. [101] my comparison of TABLE II and TABLE IV task lists and dispersions, arXiv:2509.17057. https://arxiv.org/pdf/2509.17057
MODELLED

102. [102] AIST 産総研 RoboManipBaselines repository (doc/dataset_list.md, README) and arXiv:2509.17057. https://github.com/isri-aist/RoboManipBaselines/blob/master/doc/dataset_list.md
SELF-PUBLISHED

103. [103] 株式会社レプトリノ (Leptrino) 6軸力覚センサ 製品ページ, retrieved 2026-08-12. https://www.leptrino.co.jp/product/6axis-force-sensor
SELF-PUBLISHED

104. [104] XELA Robotics uSPa 21 product page; 高密度3軸触覚センサ uSkin (カナデン 製品ページ, 検出感度の分解能 0.1グラム重); 特設ページ (368 sensing points). https://xelarobotics.com/products/uspa-21/
SELF-PUBLISHED

105. [105] 山口明彦「FingerVisionによる物体操作」日本ロボット学会誌 Vol.40 No.5, pp.399-405, 2022年6月 (解説). Cost figure from 東北大学スタートアップ事業化センター interview and Canon Foundation report (reported, not from the journal article). https://www.jstage.jst.go.jp/article/jrsj/40/5/40_40_399/_pdf
VERIFIED

106. [106] 厚生労働省 通達「産業用ロボットに係る労働安全衛生規則第150条の4の施行通達の一部改正について」平成25年12月24日 基発第1224002号 / 基発1224第2号; 日本ロボット工業会 (JARA) 産業用ロボット関連法規(抜粋). https://www.mhlw.go.jp/web/t_doc?dataId=00tb9779&dataType=1&pageNo=1
VERIFIED

107. [107] 厚生労働省 資料 ISO 10218-1:2011（JIS B8433-1:2015）ロボット及びロボティックデバイス, 項目7-8 and 力・動力の抑制 section. https://www.mhlw.go.jp/content/11300000/001367753.pdf
VERIFIED

108. [108] JIS B 8445:2016 ロボット及びロボティックデバイスー生活支援ロボットの安全要求事項 (full text via kikakurui); 日本主導で策定され、つくば市に世界初の生活支援ロボット安全検証センターが設置された. https://kikakurui.com/b8/B8445-2016-01.html
VERIFIED

109. [109] 「不整地環境における6足ロボットのコマンド指令付き歩行学習に向けたデータ拡張手法の検証」日本ロボット学会誌 Vol.44 No.5, pp.524-527, 2026年6月 (JST SPRING JPMJSP2175). https://www.jstage.jst.go.jp/article/jrsj/44/5/44_44_524/_pdf
VERIFIED

110. [110] 「速度補正を導入した高速遠隔操縦向けヒューマノイドロボット操縦システム」日本ロボット学会誌 Vol.44 No.5, pp.532-535, 2026年6月 (科研費 24K15129). https://www.jstage.jst.go.jp/article/jrsj/44/5/44_44_532/_pdf

VERIFIED

111. [111] 日本ロボット学会誌 40巻9号 (2022) 特集; 日本ロボット学会誌 44巻5号 (2026) 特集について. https://www.jstage.jst.go.jp/article/jrsj/40/9/40_40_790/_article/-char/ja
VERIFIED

112. [112] J-STAGE 全文検索 (globalSearchKey=Sim-to-Real 実機), 日本ロボット学会誌 誌面情報, retrieved 2026-08-12. https://www.jstage.jst.go.jp/browse/jrsj/-char/ja
VERIFIED

Research and review preprint, not investment advice. Figures carry the verification grade under each reference: **verified** (peer-reviewed or independently replicated), **reported** (a named source stated it, not independently confirmed), **self-published** (the organisation's own figure) and **modelled** (computed here, not measured). A modelled figure is never presented as a measurement. Corrections are welcome and will be recorded.