B2A

DATA REPORT · 30 OCTOBER 2026

AI MODEL PERFORMANCE

What has been published today about the physical AI foundational models, based on the six main criteria.

15	6	38/90	52
MODELS	CRITERIA	VALUES PUBLISHED	STILL OWED

01 — OVERVIEW

EVERY MODEL, EVERY CRITERION

Published values only. A blank criterion is owed: the vendor or authors have not published it.

MODEL	SUCCESS	ROBUSTNESS	INSTRUCTION	RECOVERY	SPEED	INTERVENTIONS
π0.7 Physical Intelligence · closed	~90% real · B	~70% real · B	~87% real · B	owed	owed	owed
GEN-1 Generalist · closed	~99% real · C	owed	owed	owed	owed	0 / 1,800+ real · C
Gemini Robotics 2 Google DeepMind · closed	78.9% real · C	owed	owed	owed	owed	owed
π0.5 Physical Intelligence · open	56.3% real · B	85.3% sim · B	71.3% sim · B	30.8% real · B	owed	owed
π0 Physical Intelligence · open	72.3% real · B	58.5% real · B	owed	owed	owed	owed
GROOT N1.7 NVIDIA · open	97.9% sim · B	75.1% sim · B	74.3% sim · B	owed	owed	owed
Cosmos Policy NVIDIA · open	97.6% sim · B	80.5% sim · B	70.5% sim · B	owed	owed	owed
UnifoLM-VLA-0 Unitree · open	97.9% sim · B	78.8% sim · B	82.2% sim · B	owed	owed	owed
Xiaomi-Robotics-0 Xiaomi · open	98.1% sim · B	73.9% sim · B	75.7% sim · B	owed	owed	owed
Fast-WAM Academic (Yuan, Dong, Liu, Zhao) · open	96.5% sim · B	70.7% sim · B	51.2% sim · B	owed	owed	owed
Wall-X X Square Robot · open	51.3% real · B	owed	owed	20.5% real · B	owed	owed
SmolVLA Hugging Face · open	32.5% real · B	owed	owed	3.2% real · B	owed	owed
OpenVLA-OFT Stanford · open	13% real · B	6% real · B	owed	owed	owed	owed
RDT-1B Tsinghua University · open	13.5% real · B	6.3% real · B	owed	owed	owed	owed
ACT Stanford (ALOHA) · baseline, not a foundation model · open	48.3% real · B	5.5% real · B	owed	6.5% real · B	owed	owed

real real robot · **sim** simulation (grey) · **A/B/C** evidence grade — A independent with logs, B paper with method, C claim with method unpublished · **~** read from a chart

02 — REPORT CARDS

ONE CARD PER MODEL

π0.7

CLOSED · PHYSICAL INTELLIGENCE

Success	■■■■■■■■■■	~90%
Robustness	■■■■■■■■	~70%
Instruction	■■■■■■■■■	~87%
Recovery	----------	owed
Speed	----------	owed
Interventions	----------	owed

3 of 6 published · full sources →

GEN-1

CLOSED · GENERALIST

Success	▨▨▨▨▨▨▨▨▨▨	~99%
Robustness	----------	owed
Instruction	----------	owed
Recovery	----------	owed
Speed	----------	owed
Interventions	▏	0/1.8k

2 of 6 published · full sources →

Gemini Robotics 2

CLOSED · GOOGLE DEEPMIND

Success	▨▨▨▨▨▨▨▨	78.9%
Robustness	----------	owed
Instruction	----------	owed
Recovery	----------	owed
Speed	----------	owed
Interventions	----------	owed

1 of 6 published · full sources →

π0.5

OPEN WEIGHTS · PHYSICAL INTELLIGENCE

Success	■■■■■■	56.3%
Robustness	■■■■■■■■■	85.3% SIM
Instruction	■■■■■■■	71.3% SIM
Recovery	■■■	30.8%
Speed	----------	owed
Interventions	----------	owed

4 of 6 published · full sources →

π0

OPEN WEIGHTS · PHYSICAL INTELLIGENCE

Success	■■■■■■■	72.3%
Robustness	■■■■■	58.5%
Instruction	----------	owed
Recovery	----------	owed
Speed	----------	owed
Interventions	----------	owed

2 of 6 published · full sources →

GROOT N1.7

OPEN WEIGHTS · NVIDIA

Success	■■■■■■■■■■	97.9% SIM
Robustness	■■■■■■■	75.1% SIM
Instruction	■■■■■■■	74.3% SIM
Recovery	----------	owed
Speed	----------	owed
Interventions	----------	owed

3 of 6 published · full sources →

■ Real robot, method published ▨ Claim, method unpublished ▭ Simulation ------ Owed

Cosmos Policy

OPEN WEIGHTS · NVIDIA

Success		97.6% SIM
Robustness		80.5% SIM
Instruction		70.5% SIM
Recovery		owed
Speed		owed
Interventions		owed

3 of 6 published · full sources →

UnifoLM-VLA-0

OPEN WEIGHTS · UNITREE

Success		97.9% SIM
Robustness		78.8% SIM
Instruction		82.2% SIM
Recovery		owed
Speed		owed
Interventions		owed

3 of 6 published · full sources →

Xiaomi-Robotics-0

OPEN WEIGHTS · XIAOMI

Success		98.1% SIM
Robustness		73.9% SIM
Instruction		75.7% SIM
Recovery		owed
Speed		owed
Interventions		owed

3 of 6 published · full sources →

Fast-WAM

OPEN WEIGHTS · ACADEMIC (YUAN, DONG, LIU, ZHAO)

Success		96.5% SIM
Robustness		70.7% SIM
Instruction		51.2% SIM
Recovery		owed
Speed		owed
Interventions		owed

3 of 6 published · full sources →

Wall-X

OPEN WEIGHTS · X SQUARE ROBOT

Success		51.3%
Robustness		owed
Instruction		owed
Recovery		20.5%
Speed		owed
Interventions		owed

2 of 6 published · full sources →

SmolVLA

OPEN WEIGHTS · HUGGING FACE

Success		32.5%
Robustness		owed
Instruction		owed
Recovery		3.2%
Speed		owed
Interventions		owed

2 of 6 published · full sources →

Real robot, method published · Claim, method unpublished · Simulation · ------ Owed

OpenVLA-OFT

OPEN WEIGHTS · STANFORD

Success	13%
Robustness	6%
Instruction	owed
Recovery	owed
Speed	owed
Interventions	owed

2 of 6 published · full sources →

RDT-1B

OPEN WEIGHTS · TSINGHUA UNIVERSITY

Success	13.5%
Robustness	6.3%
Instruction	owed
Recovery	owed
Speed	owed
Interventions	owed

2 of 6 published · full sources →

ACT

OPEN WEIGHTS · STANFORD (ALOHA) · BASELINE, NOT A FOUNDATION MODEL

Success	48.3%
Robustness	5.5%
Instruction	owed
Recovery	6.5%
Speed	owed
Interventions	owed

3 of 6 published · full sources →

Real robot, method published · Claim, method unpublished · Simulation · Owed

03 — METHOD

THE SIX CRITERIA

CRITERION	QUESTION	MEASURED AS
Success	Does it do the trained task?	% of trials, trained setup
Robustness	Does it survive the cell drifting?	% under the perturbation set
Instruction	Does rewording the instruction break it?	% under paraphrase
Recovery	Does it fix its own mistakes?	% of injected failures recovered
Speed	Is it as fast as a trained person?	human cycle time ÷ robot cycle time
Interventions	How often does a person step in?	per 1,000 cycles

EVIDENCE GRADES

- **A** Independent, with logs
- **B** Paper or report with method (rubric or error bars)
- **C** Claim; method unpublished

RULES

- Real-robot and simulation results are never mixed.
- Every value checked against its primary source on 23 Oct 2026.
- Nothing published, nothing shown: the cell is owed.

SOURCES

1. π0.7 paper (Physical Intelligence, PDF)
2. GEN-1 launch coverage, The Robot Report
3. Gemini Robotics 2, DeepMind blog
4. SO-101 VLA benchmark, arXiv 2606.08881
5. IndustrialVLA-Bench, arXiv 2609.25562
6. π0.5 paper, arXiv 2504.16054
7. Benchmarking VLAs on ALOHA, arXiv 2511.11298
8. Isaac GR00T repository (N1.7)
9. SmolVLA paper, arXiv 2506.01844

HOW TO CITE

```
Bits 2 Atoms, "AI Model Performance", b2atoms.com/data/policy-scorecard/, values checked
23 Oct 2026.
```

B2A



BITS 2 ATOMS

TRAIN YOUR ROBOT ON DEMAND.

SEE IT LIVE
Interactive charts, search and per-model pages

SUBMIT YOUR RESULTS
A value we missed or got wrong, with its source

GET NOTIFIED
An email when a model is added or a result changes

BOOK A CALL

CUSTOMERS
Test a model on your task

INVESTORS
The thesis and the plan

JOIN THE TEAM
Founding roles

hi@b2atoms.com · LinkedIn · X