Same Policy, Different Artifacts: Paired Task Audits for On-Device Neural Control

Chanseok Kang · LG Electronics · chanseok.kang@lge.com

NeurIPS 2026 Workshop on On-Device Intelligence (ODI) · ICC Sydney, 12 December 2026 · non-archival

Paper and reviews
https://openreview.net/forum?id=gvQyNdXzRc

An "INT8" label does not name one controller. Qualify a compiled artifact by paired task outcomes for its exact deployment instance, not by its checkpoint, precision label or a local error statistic.

96.6% vs 21.4%
terminal success of two Hailo-8 builds of the **same** FetchReach policy on the **same** 1,024 starts; only the calibration rows differ

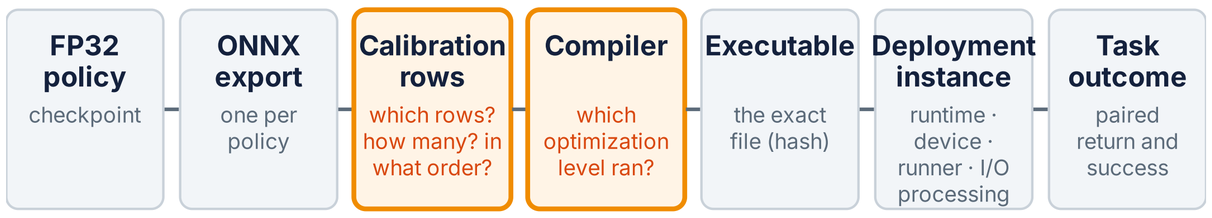
≈30 cm
best-to-worst return spread 14.93 [14.32, 15.55], as mean goal distance over the episode (Edge TPU: 0.49)

7/8
independently trained policies in which the same calibration change reproduces the gap; policy-mean 8.60 [4.42, 12.51]

1 The build record leaves out what set the controller

Post-training quantization fixes 8-bit numerical ranges from a few **calibration** states. On the audited Hailo path the compiler used its default of 64 rows and reported the count but not which rows; a CPU-only build host silently reduced optimization to level 0. Neither fact was in the record.

From checkpoint to outcome; orange boxes are what the original build record did not capture



Same checkpoint, same path, same INT8 label can still mean different controllers

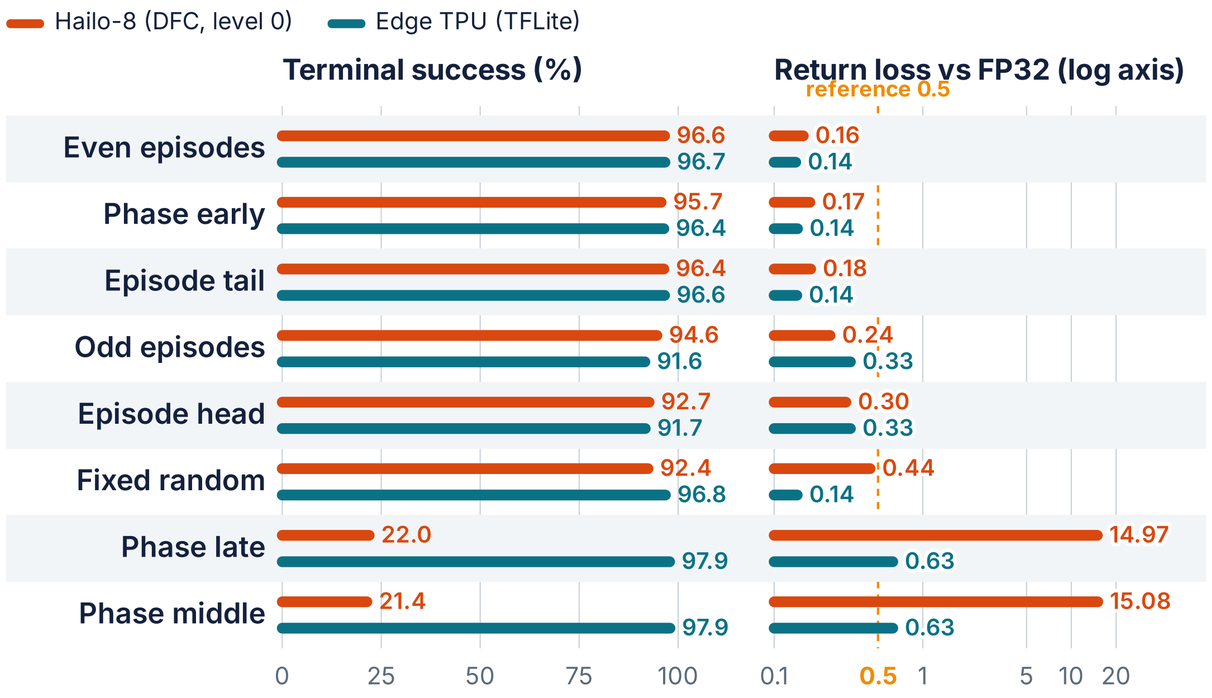
2 The audit: vary only the calibration rows

Fixed: one SAC FetchReach policy, one ONNX export, one toolchain–device path, 1,024 paired reset seeds, and the **physical accelerator in every control step**: the simulator advances only after the device returns an action.

Varied: which recorded states calibrate, by eight rules fixed before any outcome was read (even/odd episodes, episode head/tail, phase early/middle/late, fixed random).

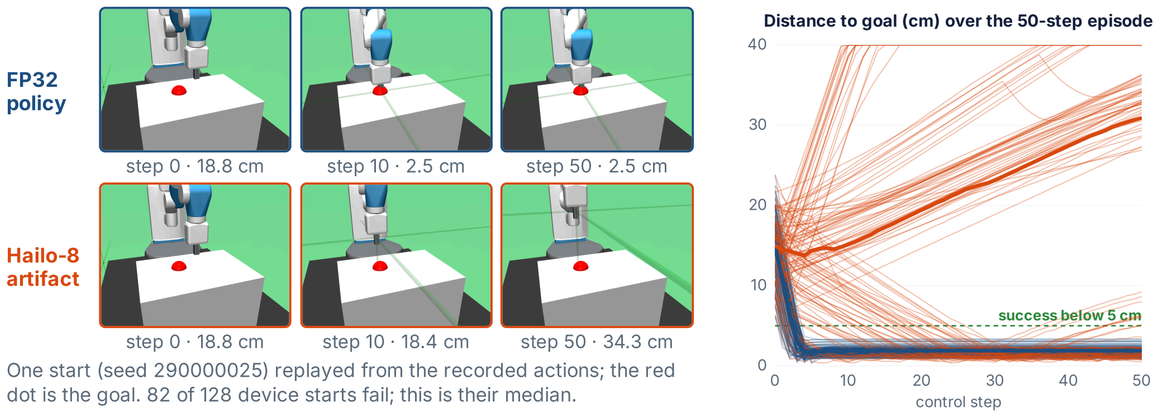
Compared: paired return (the whole route; FP32 mean -0.663) and terminal success (the endpoint) against FP32 from the same starts. 0.5 return ≈ 1 cm mean goal distance.

3 Same policy, same path, eight controllers



Eight artifacts per path, 1,024 starts each; within-path heterogeneity p < 0.001 (Holm) on both paths. On Edge TPU the two worst-return rules keep the highest success rates: return and success diverge, so either alone hides part of the behavior. Within-path effects, not a vendor comparison.

4 What a failing artifact does to the reach

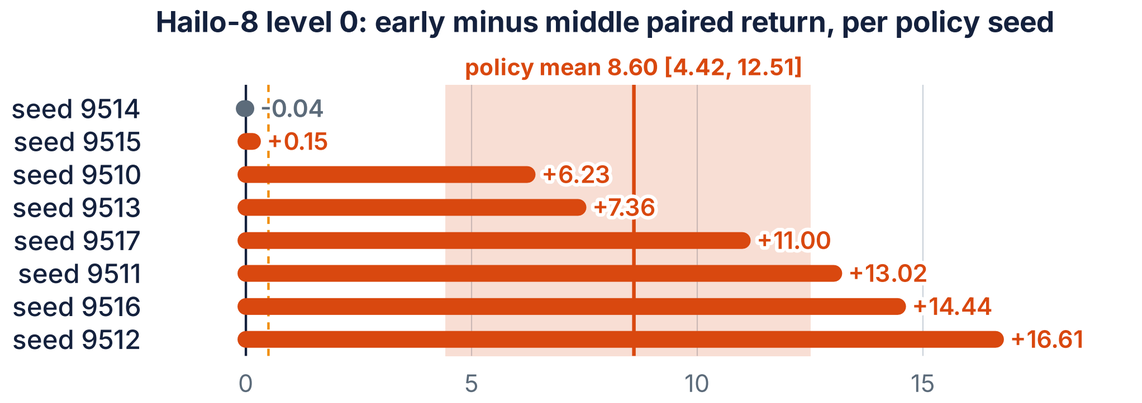


One start (seed 290000025) replayed from the recorded actions; the red dot is the goal. 82 of 128 device starts fail; this is their median.

Device-in-the-loop follow-up run on this paper's pipeline: policy seed 202609201, Hailo-8 level-0 build calibrated on 256 phase middle rows, 128 new paired starts from identical reset states. Mean return -1.06 (FP32) vs -11.07 (device). Right: every start, thin lines single starts, thick lines medians. Illustrates the mechanism; the other numbers on this poster come from the paper's cohorts.

Project page with this poster's figures, the reach as an animation and the links to paper and reviews: https://kcsgoodboy.github.io/projects/paired-task-audits-on-device/

5 It recurs across independently trained policies



Prospective replication: three rules, 256 paired starts, 8 FetchReach SAC seeds; the Hailo early-minus-middle contrast was the sole confirmatory test. Middle underperforms early in 7/8 policies (mean contrast ≈17 cm, sign-flip p = 0.0156); pooled success 45.9% vs 99.0%.

6 Forcing the compiler harder does not remove it

Hailo DFC build of the eight-rule family	Family range	95% interval	
Level 0 (auto), 64 rows, stride 4	14.93	[14.32, 15.55]	locked
Level 1 (forced), 64 rows, stride 4	14.57	[13.92, 15.21]	informed
Level 0 (forced), all 256 rows	14.97	[14.35, 15.58]	informed
Level 3 AdaRound, all 256 rows	14.42	[13.83, 15.04]	informed

Forcing level 1 changes every artifact's action field, yet the best-to-worst range stays. Neither the level-0 fallback nor AdaRound clamping alone explains the spread; "informed" rows were designed after the first result.

7 The record needs row count, identity *and* order

Supplying the same 256 rows in reverse order changed which 64 rows set the ranges: terminal success **69.9% → 97.1%** and return +1.23 [1.07, 1.39] on 512 paired starts, while the full-256 build matched an explicit first-64 control on every action and outcome. The toolchain-reported count identifies none of this.

8 What the failing artifacts change

Calibration rule	Hailo output range	FP32 outside (%)	At bound (%)	Hailo success	Edge TPU success
Episode head	[-1.01, 1.00]	1.2	0.4	**92.7**	**91.7**
Episode tail	[-1.01, 1.00]	0.0	1.2	**96.4**	**96.6**
Even episodes	[-1.01, 1.00]	0.0	1.6	**96.6**	**96.7**
Odd episodes	[-1.01, 1.00]	1.2	1.2	**94.6**	**91.6**
Phase early	[-1.01, 1.00]	0.0	1.6	**95.7**	**96.4**
Phase middle	[-0.66, 0.20]	24.2	21.1	**21.4**	**97.9**
Phase late	[-0.66, 0.20]	24.2	20.7	**22.0**	**97.9**
Fixed random	[-1.01, 1.00]	0.0	1.2	**92.4**	**96.8**

Post-outcome inspection on a fixed 800-state early-phase panel. Only the two rules with no early-phase rows calibrate the output (tanh) layer to a restricted range; a quarter of early FP32 outputs then fall outside it and Hailo-8 success collapses, while Edge TPU under the same rules stays near 98%: a contributor, not an isolated cause.

9 Local proxies do not qualify a deployment instance

5 outcome-blind artifact pairs matched within 10% in mean action-error norm and RMSE still differ by 0.08–0.17 in paired return and by up to 5.0 points in terminal success. A closed-loop amplification diagnostic (CLAE) ranked alerts better than action magnitude in selected windows but missed its pre-specified criterion on fresh data.

10 Qualify by paired outcomes, under the application's contract

> With return-loss bounds $[L_R, U_R]$ and success bounds $[L_S, U_S]$ for tolerance τ and floor ρ: **pass** if $U_R \le \tau$ and $L_S \ge \rho$; **fail** if $L_R > \tau$ or $U_S < \rho$; otherwise **indeterminate**. Families need a disjoint start block or simultaneous bounds before shipping.

Record per deployment instance: source (policy, ONNX, script hashes) · calibration (pool, rule, seed, order hashes; supplied, reported and inferred counts) · build (compiler, flags, requested and effective level, host, executable hash) · execution (runtime, device, runner, I/O processing, action-dump hash) · audit (task version, seeds, horizon, endpoint) · decision (tolerance, floor, interval, evidence status).

Takeaways

- Two executables can share checkpoint, path and INT8 label and implement different controllers.
- Calibration rows are provenance: count, identity and order all changed outcomes.
- Fixed-state error matched; task outcomes did not. Audit the complete task, paired, on the device.

Limits

- Shows that this can happen, not how often: three tasks, eight policy seeds, two accelerators.
- No GPU-host builds, real-time deadlines, physical robots or foundation-model controllers.
- The qualification rule is illustrated retrospectively, not validated end to end.

Every number on this poster is generated from the committed audit outputs; the poster manifest lists each source file and hash. Scripts, build configurations and per-run records will be released.

Build host for Hailo artifacts: CPU-only, DFC 3.33.1; Edge TPU: TFLite full-integer PTQ.