PROVENANT
AI LABS

PUBLIC SAMPLE / TECHNICAL HANDOVER

From model behaviour to an engineering decision.

Edge AI Benchmark & Technical Handover

A worked example of how Provenant frames a test, records evidence, explains uncertainty and recommends the next engineering gate.

READ THIS FIRST

Entirely synthetic. The workload, conditions, metrics and decisions in this report are fictional, created solely to demonstrate report structure. They do not describe Provenant, PeachBot or customer benchmark results.

DOCUMENT

PB2B-SAMPLE-001 | Revision 1.0

ISSUED

19 NOV 2026

Not medical advice, a clinical evaluation or a verified performance claim.

01 / DECISION BRIEF

A decision, not a dashboard.

This example begins with a business-relevant engineering question rather than an attractive but uninterpretable performance number.

ENGINEERING QUESTION

Can a quantised inference candidate meet a **150 ms p95 latency budget** and a **600 MB peak resident-memory budget** on an agreed edge-device class while preserving output agreement against the baseline?

ILLUSTRATIVE ACCEPTANCE GATES

P95 LATENCY	PEAK MEMORY	OUTPUT AGREEMENT
<= 150 ms	**<= 600 MB**	**>= 99.0%**

WORKED DECISION

CONDITIONAL GO

Proceed to an approved-data, real-device measurement phase. In the fictional example, the optimised candidate meets the nominal thresholds but misses the latency gate under thermal stress. No production release or clinical interpretation is justified.

SCOPE BOUNDARY

The example contains no real customer workload, no instrumented hardware run, no confidential source artefacts and no regulatory-grade validation. Output agreement is not equivalent to model accuracy or clinical safety.

02 / SYSTEM BOUNDARY

Make the test reproducible.

An engineering result is meaningful only when its inputs, configuration and comparison path are declared.

ILLUSTRATIVE EVALUATION PATH

01 Synthetic input

Seeded inputs; size and shape declared

02 Preprocessing

Identical transforms for both variants

03 Inference runtime

FP32 baseline and INT8 candidate

04 Evidence capture

Latency, RSS, hashes, exceptions

TEST CONTRACT

Compare like with like. Every variant uses the same input manifest, preprocessing logic, output schema and timing start/stop boundary. Capture warm-up count, measurement iterations, CPU load, thermal state, memory peak, model checksum and error count.

EXPLICITLY OUT OF SCOPE

Real patient information; sensor-integrated clinical workflows; model accuracy against trusted ground truth; embedded power integrity; hospital interoperability; resilience under field conditions; clinical safety, certification or deployment readiness.

03 / EVALUATION PROTOCOL

Describe how evidence is made.

A client should be able to challenge the measurement approach and identify what would change the conclusion.

WORKLOAD
Synthetic tensor batch

DEVICE
ARM64-class, illustrative

MODELS
FP32 vs. INT8

TRIALS
Proposed: 5 x 200

MEASUREMENT PROCEDURE

01 Freeze inputs
Record random seed, input manifest, preprocess hashes and model versions.

02 Measure baseline
Pin runtime settings; capture FP32 latency and peak-memory traces.

03 Measure candidate
Run INT8 candidate against identical inputs with identical instrumentation.

04 Apply stress condition
Repeat with defined thermal and resource constraints; capture environmental telemetry.

05 Evaluate gates
Use p95 latency, peak RSS and agreement; list all failures and missing measurements.

REPRODUCIBILITY NOTE

This document shows the proposed **shape of an evidence pack**. No test execution is represented. In a commissioned engagement, the exact device, dataset, repeat count, raw log links and benchmark scripts would be inserted before sign-off.

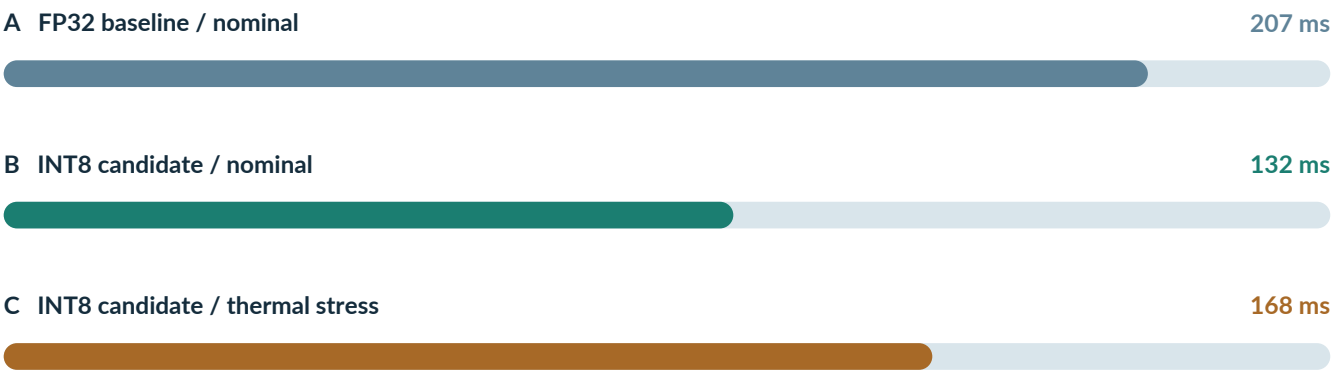
04 / ILLUSTRATIVE EVIDENCE

Read the trade-off, not one metric.

All figures on this page are fabricated presentation values. They are not measurements collected by Provenant or PeachBot.

FICTIONAL VALUES / NOT MEASURED

P95 INFERENCE LATENCY - SYNTHETIC EXAMPLE



Acceptance gate: **p95 latency <= 150 ms**. In the fictional example, B meets this limit under nominal conditions; C exceeds it under stress.

EXAMPLE RESULTS TABLE

Condition	p50	p95	Peak RSS	Agreement
A FP32 nominal	149	207	756 MB	100.0%
B INT8 nominal	94	132	525 MB	99.3%
C INT8 stress	118	168	528 MB	99.3%

Interpretation: The candidate appears promising in the fictional nominal case, but thermal stress exposes a possible deployment risk. Synthetic output agreement does not establish predictive correctness, fairness, clinical suitability or robustness.

05 / DECISION RECORD

What would we do next?

The value of the handover is a justified next decision: proceed, revise or stop, with the conditions attached.

GATE ASSESSMENT - FICTIONAL

Nominal p95 latency	132 ms <= 150 ms	PASS
Peak resident memory	525 MB <= 600 MB	PASS
Output agreement	99.3% >= 99.0%	PASS
Thermal-stress p95 latency	168 ms > 150 ms	FAIL
Real-device evidence	Required before any deployment decision	NOT TESTED

RISK AND MITIGATION

Thermal headroom	Profile sustained inference and throttling on the agreed hardware.
Quality comparability	Replace synthetic agreement with approved-data task metrics and stratified error review.
Instrumentation drift	Version scripts, model hashes, runtime flags and raw trace checksums.

RECOMMENDATION

Conditional GO to the next measurement phase. Require signed-off device specifications, legitimate dataset access, agreed workload and stop conditions. A failed thermal gate blocks any claim of readiness.

06 / HANDOVER REGISTER

Leave an audit trail.

A completed paid engagement should provide an independently reviewable chain from question to decision.

ILLUSTRATIVE HANDOVER PACKAGE

01 **Scope & acceptance record**
Signed SOW; agreed question, thresholds, exclusions and data controls.

02 **Evaluation manifest**
Model/runtime identifiers, input manifest, device settings and versions.

03 **Raw evidence package**
Timestamped per-run telemetry, error logs and checksum index.

04 **Analysis workbook**
Metric definitions, aggregation procedure and comparison outputs.

05 **Engineering decision brief**
Gate assessment, risks, limitations and approved next step.

CLIENT HANDOVER CHECKLIST

Confirm before sign-off: the question and test gates are unchanged; all numerical claims link to raw evidence; deviations are recorded; IP and licensing terms are known; data retention is agreed; open risks have owners; and the decision-maker acknowledges residual uncertainty.

EDITORIAL REFERENCE

This sample follows the documentation discipline of the **PeachBot EcoSense Technical Whitepaper** (architecture, methodology, evaluation, governance and limitations). EcoSense content and field results have **not** been repurposed as benchmark evidence. Public reference: peachbot.in/assets/pdf/PeachBot EcoSense.pdf.