The Steel.vision Evaluation Framework

Independent evaluation for AI in environments where failure is expensive.

A fixed-price, two-week evaluation sprint ending in the evidence pack your risk committee needs.

WHAT WE TEST

- **Reliability & hallucination**
  Does the system assert falsehoods with confidence — and under what operating conditions?
- **Robustness & gaming**
  Adversarial probing: prompt manipulation, metric gaming, and behaviour at the edges of the training distribution.
- **Regression & drift**
  Behaviour versus baseline across versions — what silently changed, and whether monitoring would catch it.
- **Governance evidence**
  The documented trail regulators and risk committees expect: what was tested, how, by whom, with what result.

WHY OUR EVALUATIONS HOLD UP

- **Proven where money is real**
  Our validation stack runs daily on a live, real-money machine-learning trading system: purged cross-validation, sealed holdouts, drift monitoring, fail-closed deployment gates.
- **Grounded in published research**
  Four papers on machine-learning evaluation — including how models game the very metrics used to judge them.
- **Built, not just audited**
  Our founder ships production AI, including a safety-critical, fully offline Edge AI system for industrial field crews. We evaluate AI where it actually runs — cloud, on-premise, or air-gapped on-device.

WHAT YOU RECEIVE

- Findings report, ranked by severity
- Test methodology & coverage map
- Reproducible evaluation-harness outputs
- A governance pack written for the committee that signs — not the engineers who already believe

THE ENGAGEMENT & WHAT IT COSTS

Two weeks, kickoff to evidence pack.

Fixed price, agreed in writing before kickoff — one number, no day-rate drift.

Scope set in a 30-minute call: what your system does, where it runs, what your committee needs to see.

START A CONVERSATION → steel.vision/assurance

Steel.vision is a trading name of Steel Vision Ltd · Registered in Scotland · Company No. SC901577
Edinburgh, United Kingdom · steel.vision