

Construction AI Evaluation Scorecard

Use one copy per product and workflow. Score the accepted operating outcome, not the quality of a sales demonstration.

Company	
Workflow being evaluated	
Product and exact plan/version	
Business owner	
Technical owner	
Evaluation period	
Decision deadline	

Scoring

0 = not evidenced · 1 = material gap · 2 = partly evidenced · 3 = acceptable for pilot · 4 = strong evidence in representative cases · 5 = accepted in the intended operating boundary

Criterion	What to verify	0-5	Evidence / gap / owner
Workflow fit	The product supports the exact trigger, inputs, users, stages, output, and completion condition.		
Source identity	Every input retains project, record, document, revision, date, owner, and authoritative source.		
Evidence and citations	Prepared outputs link to the exact page, drawing, image, schedule, record, or field used.		
Permissions	User, service, project, record, field, tool, action, and destination access are least-privilege and testable.		
Integration behavior	Read/write direction, native IDs, duplicates, idempotency, failures, retries, and reconciliation are controlled.		
Human authority	Technical, safety, commercial, contractual, payment, personnel, and external actions remain explicitly authorized.		

Criterion	What to verify	0-5	Evidence / gap / owner
Critical-error behavior	The product refuses, flags, or escalates missing, conflicting, ambiguous, restricted, and unsupported cases.		
Representative evaluation	Normal, difficult, incomplete, conflicting, restricted, edge, and failed cases have expected outcomes.		
Monitoring and recovery	Runs, versions, changes, incidents, alerts, correction, rollback, and recovery have named owners.		
Complete cost	Licence, implementation, integration, data, compute, review, correction, support, training, and exit are included.		
Ownership and portability	Client records, outputs, corrections, configurations, documentation, and accepted state can be exported and reused.		
Provider exit	There is a tested path for contract end, price change, product retirement, model replacement, and data deletion.		

Pilot acceptance record

Accepted workflow boundary

Trigger and completion condition	
Users and responsible owner	
Approved sources and systems	
Permitted AI outputs and actions	
Prohibited or consequential authority	

Evaluation evidence

Representative case count	
Critical failure count	
Accepted outcome rate	
Average reviewer correction time	
Complete cost per accepted outcome	

Operating readiness

Monitoring and alert owner	
Exception and recovery owner	
Security and incident owner	
Training and adoption owner	
Provider change and exit owner	

Decision

- ☐ Stop — no suitable workflow or unacceptable critical boundary
- ☐ Revise — evidence or controls are incomplete
- ☐ Pilot further — expand representative cases without production promotion
- ☐ Proceed with bounded implementation — written scope and acceptance required
- ☐ Monitor — relevant capability, not ready for this workflow

Decision owner / date / conditions

Important boundary

This scorecard supports product evaluation. It does not certify product accuracy, cybersecurity, privacy, legal compliance, professional suitability, safety performance, commercial value, or production readiness. Verify current product capabilities, contracts, region, data use, pricing, integrations, and support directly with the vendor.