Verdify
AI CONSULTING

FLAGSHIP BLUEPRINT

Verdify AI Control Loop Blueprint

Moving from AI pilot to Verified AI workflow

The promise

A practical executive guide for putting AI into real work without losing action limits, human approval, system authority, telemetry, or evidence.

Map | Define Controls | Build | Verify | Operate

Verdify Lab is the public proof environment. The PDF applies the same operating pattern to business workflows.

www.verdify.ai

THE MOVE

From plausible answer to operating workflow.

Most pilots prove that a model can respond. They do not prove that the response can enter a live workflow in a way leadership can govern, operators can trust, and technical teams can monitor. The missing artifact is an operating loop.

1. Context before model

The workflow, owner, source systems, exception paths, and failure modes are named before a model receives authority.

2. Controls before writes

The AI agent may classify, draft, recommend, or prepare evidence. Risky writes pass through control layers first.

3. Evidence before expansion

Telemetry, reviewer signal, overrides, incidents, and outcomes decide whether scope expands, holds, or stops.

What Verdify means by Verified AI

Verified AI is not a claim that a model is always right. It is a way of operating AI where the role is explicit, the action surface is narrow, the authority layer is preserved, and the workflow produces evidence that leaders can inspect.

External operating pattern

The blueprint aligns with recognized production practice: contextual risk management, simple composable workflows, output validation before downstream action, least privilege, explicit authorization, post-deployment monitoring, and change control.

THE PRACTICAL FAILURE MODE

Pilots stall when the operating system is missing.

The problem is rarely that the model cannot produce a useful answer. The problem is that nobody has defined where that answer may go, who can approve it, what evidence supports it, or how the workflow will be reviewed after launch.

Unmapped workflow

The trigger, owner, system of record, exception path, and failure mode are still vague.

Too much action surface

The model sees too many tools, fields, permissions, or implicit decisions for the risk involved.

No control layer

Output is not checked for structure, evidence, permission, freshness, or approval before downstream use.

No operating scorecard

The team cannot show override rate, incidents, rework, cycle time, quality, or outcome movement.

The control loop fixes the handoff problem.

Context is assembled. Intent is constrained. Outputs are checked before downstream writes. Accepted setpoints move through the right authority layer. Operators can intervene. Outcomes feed the next review cycle.

COMMERCIAL METHOD

Map / Define Controls / Build / Verify / Operate

The method is intentionally plain. It gives executives, operators, and technical teams one shared sequence for deciding whether an AI workflow deserves production scope.

1

Map

Name the workflow, owner, trigger, systems, evidence, failure modes, and current baseline.

2

Define Controls

Set the action limits, approvals, authority layer, prohibited actions, rollback path, and known limits.

3

Build

Implement the narrowest useful workflow with structured context, routing, checks, and controlled integrations.

4

Verify

Test the workflow in deployment-like conditions and review quality, exceptions, telemetry, and reviewer signal.

5

Operate

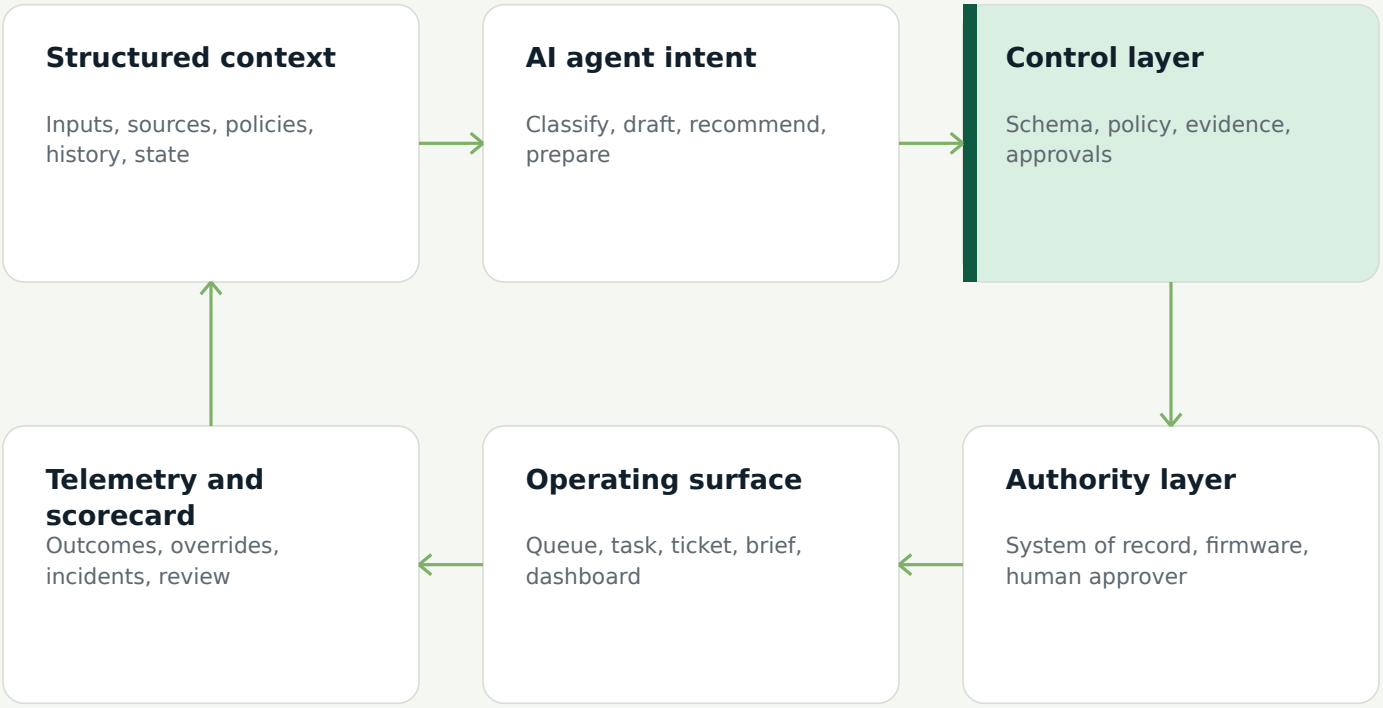
Run the workflow with scorecards, overrides, incident review, known limits, and evidence-based expansion gates.

The sequence is not a maturity slogan. It is a release discipline: scope earns authority only after the evidence is strong enough for the next operating decision.

ARCHITECTURE

The loop separates reasoning, validation, authority, and evidence.

A useful AI workflow is a series of disciplined handoffs. The model may reason and propose, but the control layer decides whether an output is admissible, and the authority layer decides what can affect the real workflow.



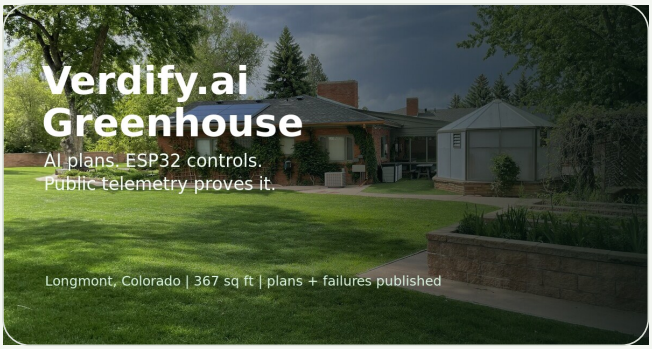
Operating rule

The AI proposes. The control layer checks. The authority layer decides. The operator can intervene. The scorecard determines whether the workflow earns more scope.

PROOF PATTERN

The greenhouse makes Verified AI inspectable.

Verdify Lab is a real operating environment: weather changes, sensors drift, hardware has limits, and control decisions matter. That makes it useful as public proof for a broader business pattern.



Lab bridge

The AI agent plans. Control layers constrain writes. Firmware controls. Telemetry verifies. Scorecards and lessons close the loop.

AI planning role

Reads operating context and proposes tactics through approved fields.

Control layer

Checks structure, ranges, required IDs, ownership, and unsupported writes.

Authority layer

Firmware owns local physical control and deterministic safety behavior.

Evidence layer

Telemetry, delivery logs, readbacks, caveats, scorecards, and lessons make outcomes inspectable.

The point is not greenhouse automation. The point is separation of concerns: model influence is constrained, execution authority is preserved, and evidence remains visible.

WORKED EXAMPLE

Cleantech pilot-to-procurement evidence pack

Strong pilots can still die in procurement when assumptions, evidence, risk posture, and diligence answers are scattered. The control loop turns pilot artifacts into a buyer-ready evidence package without letting AI overstate the result.

Loop element	What AI may do	What remains controlled
Structured context	Consolidate KPIs, deployment notes, assumptions, safety notes, and buyer questions.	Approved source artifacts only.
Control layer	Map claims to evidence and flag unsupported assertions.	No invented evidence or hidden assumptions.
Authority layer	Prepare diligence responses and risk notes.	Humans approve performance claims and customer-facing language.
Telemetry	Track evidence gaps, reviewer overrides, cycle time, and accepted packet sections.	The scorecard determines whether the workflow expands.

Claim limit

AI may not certify savings, sign contracts, claim customer endorsement, approve commercial commitments, or submit buyer-facing materials without human approval.

PATTERN PORTABILITY

The same loop works when the workflow is not physical.

A business workflow usually does not have firmware or relays. It does have records, approvals, policies, queues, deadlines, customer impact, and consequences. The architecture transfers when those elements are made explicit.

Lab context	Business equivalent	Example: invoice exception handling
Structured context	Case packet	Invoice, PO, receipt, vendor terms, exception history, policy, SLA
AI agent intent	Tactical workflow intent	Classify mismatch, explain likely cause, recommend next step
Control layer	Validation gate	Check schema, policy, evidence, confidence, and approval threshold
Authority layer	System-of-record authority	ERP, approval workflow, and finance controls retain execution authority
Operator brief	Human-facing work summary	Show evidence, uncertainty, recommendation, and required approval
Telemetry	Operating review	Resolved, overridden, escalated, delayed, rejected, or reopened cases

Safe translation: the AI proposes, the control layer checks, the authority layer decides, the operator can intervene, and the scorecard tells leadership whether broader scope is earned.

PRODUCTION CONTROLS

Downstream writes need a gate, not just a prompt.

The validation layer is where pilot behavior becomes production behavior. Every proposed write should answer five questions before it can affect a workflow.

1. Is the output structurally valid?
2. Is the requested action allowed in this workflow?
3. Is the supporting evidence present and current?
4. Is the action inside the approved risk threshold?
5. Is the final authority for execution correctly identified?

Control layer

Schemas, policy rules, retrieval checks, confidence thresholds, reviewer gates, prohibited-action tests, and exception routing.

Authority layer

The ERP, CRM, workflow engine, firmware, approval service, or named human approver that owns real execution.

OPERATING EVIDENCE

The scorecard is how scope gets earned.

A control loop without telemetry is just automation with better vocabulary. The workflow needs a record of what happened, who approved it, what changed, and whether outcomes improved.

Flow efficiency

Turnaround time, backlog age, manual touches, first-pass completeness, queue aging.

Control health

Missing-source rate, unsupported-claim rate, override rate, exception recurrence, incident flags.

Business outcome

Deal speed, release safety, audit findings, rebate lag, quality recurrence, resolution quality.

Minimum telemetry baseline

- workflow/request ID
- workflow configuration version
- validation result and exception reason
- override events and timestamps
- source records and revisions
- current control rules and approval requirements
- human reviewer and final approver
- traceable final output and approval record

RELEASE DISCIPLINE

Scale evidence before scope.

A larger action surface should be earned by stable behavior, clean exception handling, and a scorecard leadership trusts. Do not expand because the demo looked good. Expand because the operating evidence is good.

1. Shadow mode
AI produces plans and briefs. Nothing is written downstream.

→

2. Assisted mode
Operators accept or reject controlled setpoints one case at a time.

→

3. Limited authority
Low-risk writes proceed only after strong validation and review evidence.

→

4. Controlled expansion
Scope expands when override, incident, and outcome signals stay inside agreed limits.

Expansion gates

Expand when telemetry is clean, reviewer trust is high, exceptions are understood, and the next authority step has a named owner.

Hold or stop gates

Hold or stop when unsupported claims, overrides, incidents, stale sources, or unclear authority appear in the operating record.

The release decision should be explicit: build, audit first, scorecard first, hold, or stop.

USE THIS PAGE IN REVIEW

A one-page operating brief for the next AI workflow.

Workflow

What repeated work should AI help with?

AI role

Read, classify, draft, recommend, prepare evidence, or propose setpoints.

Prohibited actions

What must AI never do in the first release?

Authority layer

Which system, policy, firmware, or person owns final execution?

Control layer

Which checks run before any downstream write?

Telemetry

What record proves what happened?

Scorecard

Which metrics decide expand, tune, hold, or stop?

Known limits

What is not proven yet?

Decision

Build a controlled MVP | Run an audit first | Define a scorecard first | Hold | Stop or defer

NEXT STEP

Start with one workflow you can defend.

Verdify's Verified AI Operations Audit helps teams decide where a pilot can move into controlled production, what should remain human-owned, what telemetry is missing, and what evidence is required before release.

Verified AI Operations Audit

Map the workflow. Define the operating surface. Design the validation and authority layers. Instrument the scorecard. Leave with a practical build, hold, or stop decision.

Source basis

This blueprint synthesizes Verdify Lab's public proof pattern with recognized production guidance for AI risk management, agent workflow design, LLM application security, and tool authorization. It is not legal, regulatory, security, or compliance advice.

1. NIST AI Risk Management Framework 1.0	https://www.nist.gov/itl/ai-risk-management-framework
2. NIST AI 600-1 Generative AI Profile	https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
3. Anthropic, Building Effective Agents	https://www.anthropic.com/engineering/building-effective-agents
4. OWASP Top 10 for Large Language Model Applications	https://owasp.org/www-project-top-10-for-large-language-model-applications/
5. Model Context Protocol authorization specification	https://modelcontextprotocol.io/specification/2025-03-26/basic/authorization
6. Verdify Lab public proof environment	https://lab.verdify.ai/

Book a Fit Call or start with a Verified AI Operations Audit at www.verdify.ai.