True Function Alignment Map

A HAOP diagnostic worksheet for testing whether an AI-enabled workflow is producing the safety outcome it claims to pursue - or only the appearance of safety.

> **Use this before AI output becomes consequential.**
> The map surfaces misalignment across human judgment, AI optimization, and organizational signals before drift becomes embedded in the workflow.

Purpose

The worksheet converts the True Function Test into a practical map. It asks whether the workflow remains grounded in operational reality, whether control is assigned to the people or functions that actually hold it, and whether verification and pause points exist before AI-enabled outputs shape action.

How to use it

1. Pick one workflow. Select a specific workflow where AI classifies, routes, scores, recommends, approves, generates controls, monitors compliance, or shapes what a human sees first.

2. Answer the nine diagnostic questions. Identify the declared purpose, actual signal, judgment relied on, behavior allowed, verification point, and pause path.

3. Map across all three performers. Fill in what is happening for the human performer, AI performer, and organizational performer.

4. Look for misalignment. Risk increases when the declared purpose is safety but the live signal is speed, compliance appearance, volume, cost reduction, avoidance of bad news, or friction removal.

5. Convert findings into design actions. Add constraints, verification points, escalation triggers, pause authority, monitoring, auditability, or ownership changes before the workflow scales.

Core HAOP premise

Humans adapt. AI optimizes. Organizations signal. The worksheet forces those three forms of performance into the same view, instead of letting accountability collapse onto the person closest to the failure.

Expected output

The output is not a score. It is a clearer view of where the workflow is aligned, where it is drifting, and what must be redesigned before the system scales.

Part 1: The Nine Diagnostic Questions

Answer these before completing the worksheet. The value comes from forcing the answers across human, AI, and organizational control points.

Q1. What outcome is claimed?	What safety, quality, operational, or compliance outcome does the workflow say it is producing?
Q2. What signal is optimized?	What metric, target, proxy, prompt, score, dashboard, or incentive is actually shaping the workflow?
Q3. What judgment is relied on?	What must the human notice, interpret, challenge, verify, or decide, and what is the AI asked to classify, recommend, generate, approve, or route?
Q4. What behavior is allowed?	What shortcuts, adaptations, overrides, permissions, autonomous actions, or tolerated exceptions can occur without review?
Q5. Where does verification occur?	Where is output checked against real conditions, known constraints, source data, operational limits, or field expertise before it becomes consequential?
Q6. Who can stop the workflow?	Can a worker, reviewer, supervisor, technical owner, or governance function pause, challenge, override, or escalate the process? Is that authority protected and resourced?
Q7. What can early failure look like?	What weak or low-noise indicators would show that the workflow is drifting before harm occurs?
Q8. What is shaping behavior?	What context, fatigue, training, peer norms, data limits, model limits, permissions, incentives, policies, budgets, or priorities are pushing the system?
Q9. What weak signals may be erased?	What tacit knowledge, edge cases, anomalies, dissent, bottlenecks, informal warnings, near misses, missing context, or underreported risk may disappear as the workflow becomes cleaner and faster?

Part 2: Complete the Alignment Map

Use the worksheet on the next page to map each question across the human performer, AI performer, and organizational performer. The final row asks what control is needed.

Red flags

Declared purpose is safety but the active signal is speed, cost, volume, closure rate, or appearance of compliance; Oversight exists as a label but the reviewer lacks time, skill, authority, source data, or protected pause authority; AI can route, suppress, escalate, approve, or trigger action without a defined verification point; No owner is assigned for constraints, monitoring, exceptions, drift, or escalation; The workflow looks cleaner while operational reality becomes less visible.

Appendix B: True Function Alignment Map

A free HAOP worksheet for testing whether an AI-enabled workflow is producing the safety outcome it claims to pursue — or only the appearance of safety.

Run each question across all three performers:

Map Section	Prompt	Human Performer	AI Performer	Organizational Performer
Framing Prompt	**What work is being performed?**	judgment, adaptation, review, escalation	classification, generation, routing, optimization	work design, approval, resourcing, metrics
Q1	**What outcome is claimed?**	What is the person expected to accomplish?	What is the AI expected to produce or influence?	What business or safety outcome is the organization claiming?
Q2	**What signal is optimized?**	What behavior is rewarded, pressured, or normalized?	What data, metric, prompt, target, or proxy is the AI optimizing?	What metric, incentive, or executive priority is shaping the workflow?
Q3	**What judgment is relied on?**	What must the human notice, interpret, challenge, or verify?	What does the AI classify, recommend, generate, approve, or route?	What assumptions has the organization made about human capacity and AI reliability?
Q4	**What behavior is allowed?**	What shortcut, adaptations, or overrides are tolerated?	What can the AI do without review or intervention?	What permissions, resources, and authority has the organization granted?
Q5	**Where does verification occur?**	Does the human have time, skill, context, and authority to verify?	Is the AI output checked against reality, constraints, or known limits?	Has the organization defined points of consequence and required validation there?
Q6	**Who can stop the workflow?**	Can the worker or reviewer pause or challenge the process?	Are there technical limits, hard stops, or escalation triggers?	Has stop-work authority been designed, protected, and resourced?
Q7	**What can early failure look like?**	Fatigue, silence, workarounds, attention or judgment miss, compliance miss, overtrust	Confident incompetence, false negatives, signal substitution, constraint failure	Dashboard confidence, compliance theater, accountability gaps, normalized deviance, weak oversight
Q8	**What is shaping behavior?**	context, fatigue, training, peer norms	data, prompts, model limits, permissions	incentives, policies, budgets, priorities
Q9	**What weak signals may be erased?**	Hesitation, informal warnings, tacit knowledge, near misses	Edge cases, anomalies, uncertainty, missing context	Bad-news resistance, bottlenecks, friction, dissent, underreported risk, resource constraints
Control Synthesis	**What control is needed?**	time, authority, training, stop-work ability	validation, constraints, monitoring, auditability	governance, ownership, escalation, consequence design