

THE HAPI AUDIT MODEL

Measuring Agency Preservation Across Systems and Institutions

A Human Agency Preservation Infrastructure Working Thesis Paper

Author: Michael Bower

Project: HAPI, Human Agency Preservation Infrastructure

Version: v0.1

Date: May 18, 2026

Status: Working research manuscript, not peer reviewed

Core thesis: A system can be audited for whether it preserves, weakens, captures, or restores meaningful human agency.

Abstract

This paper formalizes the HAPI Audit Model, a diagnostic method for evaluating whether a system preserves meaningful human agency. The model begins from the claim that agency is not merely the presence of choice, consent, approval, or participation language. Agency exists when humans can understand what is happening, exercise rightful authority, refuse before consequence, revise proposed actions, remember what occurred, contest outcomes, and remain accountable for consequences they could meaningfully affect. Modern systems often preserve the appearance of human participation while weakening the conditions that make participation real. This produces agency theater: humans remain visible in the system, but their power to understand, refuse, revise, or affect outcomes has been reduced. The HAPI Audit Model provides a practical structure for detecting this failure across institutions, AI systems, workplaces, religious communities, healthcare pathways, governance processes, and everyday administrative systems. It evaluates clarity, authority, refusal, revision, memory, contestability, capacity, accountability, timing, dependency risk, and policy-reality coherence. The purpose of the audit is not punishment, surveillance, or abstract scoring. Its purpose is restoration. A HAPI audit identifies where agency has been preserved, weakened, captured, or restored, then produces a restoration map that helps the system rebuild governance from the core problem outward. The central conclusion is that governance cannot be validated by policy artifacts alone. Governance becomes real when agency remains live before, during, and after consequential system action.

Keywords

human agency, agency preservation, agency loss, governance, audit model, meaningful participation, refusal, accountability, contestability, agency theater, HAPI

Table of Contents

1. Introduction
 2. The Purpose of a HAPI Audit
 3. Core Audit Question
 4. Agency States: Preserved, Weakened, Captured, Restored
 5. Audit Dimensions
 6. The Agency Chain Map
 7. The Consequence Boundary
 8. Evidence Types
 9. Finding Types and Severity
 10. Restoration Mapping
 11. Continuity and Re-Audit
 12. Applications Across Domains
 13. Limits and Failure Modes
 14. Conclusion
- Appendix A: Audit Worksheet
- Appendix B: Finding Template
- Appendix C: Glossary

1. Introduction

HAPI begins with a simple but demanding claim: governance is the product of agency preservation. A system cannot be considered well governed merely because it has policies, approvals, dashboards, procedures, or compliance language. These artifacts may be useful, but they do not prove that humans retain meaningful participation at the point where consequences are created.

The HAPI Audit Model exists to test whether agency is actually present. It asks whether the relevant human beings can understand, refuse, revise, act, remember, contest, and remain accountable. It also asks whether the system gives responsibility to people who lack authority, gives authority to actors who lack accountability, or uses human presence as a symbol while decisions move elsewhere.

This is especially important in systems shaped by automation, institutional pressure, role hierarchy, spiritual authority, administrative complexity, and AI delegation. In these environments, people may technically remain in the loop while losing the ability to affect the outcome. The audit model gives HAPI a repeatable method for finding those agency failures and turning them into restoration work.

Core audit principle

A HAPI audit does not ask only whether a human was present. It asks whether human agency was still live.

2. The Purpose of a HAPI Audit

The purpose of a HAPI audit is to identify where agency is preserved, weakened, captured, or restored inside a system. The audit is not designed to shame individuals or produce a simplistic score. Its purpose is to reveal the structural conditions that either support or destroy meaningful human participation.

A HAPI audit has five practical goals:

1. **Locate agency loss.** Identify where people are present but unable to meaningfully understand, refuse, revise, or affect outcomes.
2. **Separate real governance from theater.** Distinguish actual authority from symbolic approval, consent language, or procedural formality.
3. **Map responsibility to authority.** Find places where people are held responsible for consequences they could not direct.
4. **Preserve memory.** Determine whether the system can remember what happened, why it happened, who approved it, and what evidence was available.
5. **Build a restoration path.** Recommend changes that restore clarity, refusal, revision, memory, contestability, and accountability.

The audit therefore serves the larger HAPI mission: restore agency where systems have reduced people to rubber stamps.

3. Core Audit Question

Every HAPI audit begins with one central question:

Can the affected human or group meaningfully participate before, during, and after consequential system action?

This question can be expanded into three phases.

Before action

- Can the human understand what is proposed?
- Can the human identify who or what has authority?
- Can the human refuse before consequence?
- Can the human revise the proposed action?
- Can the human see the evidence and assumptions shaping the decision?

During action

- Does execution remain bound to what was authorized?
- Can the human halt or escalate if scope changes?
- Does the system prevent drift from proposal to action?
- Does the system preserve the difference between support and substitution?

After action

- Can the human see what happened and why?
- Can the human contest or appeal the result?
- Can the system preserve memory for accountability and learning?
- Does accountability map to the people or entities that actually had control?

4. Agency States: Preserved, Weakened, Captured, Restored

The audit model classifies agency conditions into four states. These states are not moral labels placed on people. They are diagnostic categories for understanding the structure of the system.

State	Meaning
Preserved	Humans retain clarity, authority, refusal, revision, memory, and accountability at the relevant consequence boundary.
Weakened	Some conditions of agency remain, but pressure, confusion, missing context, delay, overload, or weak documentation reduces meaningful participation.
Captured	The system, institution, agent, or authority layer takes over decision power while preserving the appearance of human participation.
Restored	Previously weakened or captured agency has been rebuilt through better structure, boundaries, memory, authority, and support.

The most important distinction is between weakened agency and captured agency. Weakened agency may occur because the system is poorly designed, confusing, overloaded, or underdeveloped. Captured agency occurs when another actor or structure benefits from the human losing meaningful participation while still carrying the language or burden of responsibility.

5. Audit Dimensions

The HAPI Audit Model evaluates eleven agency dimensions. These dimensions are the load-bearing conditions that make participation real.

- **Clarity:** Can the human understand what is happening, what is proposed, and what the consequences may be?
- **Authority:** Does the human or role have rightful power to approve, refuse, revise, escalate, or halt the action?
- **Refusal:** Can the human say no before consequence without being punished, bypassed, or coerced?
- **Revision:** Can the human change the proposal before it becomes committed action?
- **Memory:** Can the system preserve what happened, why, who decided, what evidence existed, and what changed?
- **Contestability:** Can the human challenge, appeal, correct, or re-open an outcome?
- **Capacity:** Does the human have the time, energy, knowledge, and support needed to participate meaningfully?
- **Accountability:** Is responsibility assigned to the actor who had real authority and control?
- **Timing:** Does human participation happen early enough to affect the outcome?
- **Dependency risk:** Does the support system restore human ability, or does it make the human increasingly unable to act without it?
- **Policy-reality coherence:** Do written policies match the actual operating behavior of the system?

Audit warning

If any one of the core conditions approaches zero, agency can collapse even when the system still appears procedurally complete.

6. The Agency Chain Map

Before findings can be written, the auditor must map the agency chain. This identifies who has intent, who proposes action, who reviews it, who authorizes it, who executes it, who is affected, and who carries accountability afterward.

A basic HAPI agency chain looks like this:

Human or organizational intent -> proposal -> review -> authorization -> execution -> consequence -> memory -> accountability

The audit looks for breaks in this chain. A break occurs when intent is distorted, review is symbolic, authorization is unclear, execution drifts, consequence is hidden, memory is incomplete, or accountability is assigned to the wrong actor.

In AI systems, the chain becomes more complex because the agent may transform intent into tool calls or workflow actions. In institutions, the chain becomes complex because authority may be distributed across policy, supervisors, committees, forms, social pressure, and informal norms. In religious or moral communities, the chain may involve conscience, spiritual authority, belonging, family relationships, and claims about divine approval.

7. The Consequence Boundary

A central HAPI concept is the consequence boundary. This is the moment when a proposal becomes action, a judgment becomes status, an approval becomes commitment, a workflow updates reality, or a decision begins affecting people.

Many systems preserve human agency before discussion but remove it at the consequence boundary. A person may be informed, consulted, or asked to approve, but the actual action may already be effectively decided. The audit therefore asks whether agency is live at the boundary where the system becomes consequential.

- **Weak boundary:** The human is notified after the action has effectively happened.
- **Symbolic boundary:** The human clicks approval but cannot meaningfully understand or refuse.
- **Captured boundary:** The system requires a human signature while another actor controls the real decision.
- **True boundary:** The human can understand, refuse, revise, escalate, or halt before the action commits.

Human review is not oversight unless it can still change the outcome.

8. Evidence Types

The audit should not depend only on what a system claims about itself. It should compare written claims, actual workflows, user experience, system logs, receipts, interviews, and observed behavior.

Useful evidence includes:

- policy documents and stated values
- workflow diagrams and approval paths
- system logs, receipts, permits, tickets, and audit trails
- training material and role expectations
- interviews with affected humans and responsible authorities
- examples of contested decisions or failed escalations
- records of exceptions, overrides, and bypasses
- timing data showing whether review happened before or after consequence
- evidence of whether users understood the decision and had meaningful options

The strongest evidence comes from mismatch. If policy says humans can refuse but interviews show refusal is punished, the audit should treat the policy as incomplete. If a dashboard claims approval but logs show execution happened first, the audit should treat the approval as theater. If a

system preserves receipts but those receipts cannot be accessed by the affected human, memory may exist for the institution but not for agency restoration.

9. Finding Types and Severity

A HAPI finding should be written as a restoration-oriented diagnosis. It should describe what agency condition failed, who was affected, where the consequence boundary occurred, and what restoration would require.

Core finding types

- **Agency gap:** A required agency condition is missing or weak.
- **Agency theater:** The system preserves the appearance of participation while removing meaningful power.
- **Authority mismatch:** Responsibility and authority are assigned to different actors.
- **Refusal failure:** The human cannot say no before consequence, or refusal is punished or bypassed.
- **Memory failure:** The system cannot preserve what happened, why, who approved it, or what evidence existed.
- **Contestability failure:** The human cannot challenge, appeal, correct, or understand the outcome.
- **Dependency capture:** The support structure makes the human less able to act over time.
- **Policy-reality split:** Written governance claims do not match the real operating system.

Severity logic

Severity should not be based only on the size of the policy violation. It should be based on the degree of agency loss and the consequence of that loss.

- **Low:** Agency is inconvenienced or made less clear, but refusal, revision, memory, and contestability remain available.
- **Moderate:** Agency is meaningfully weakened, but the person can still recover participation with support or escalation.
- **High:** Agency is structurally captured at the consequence boundary, or the person carries responsibility without control.
- **Critical:** The system strips agency in a way that affects safety, livelihood, family, identity, legal standing, health, spiritual belonging, or irreversible consequence.

10. Restoration Mapping

A HAPI audit is incomplete if it only names failures. The audit must also define a restoration map. The restoration map identifies what conditions must be rebuilt so agency becomes real again.

Restoration recommendations should be concrete. They should not merely say “improve transparency” or “increase oversight.” They should state which agency function must be restored and how.

Examples of restoration moves

- **Restore clarity:** Rewrite decision language so affected humans can understand the proposal, risk, authority, and consequence.
- **Restore refusal:** Move review earlier in the workflow and guarantee that refusal pauses execution.
- **Restore revision:** Give reviewers the ability to return an action for correction instead of only approve or deny.
- **Restore memory:** Create receipts that preserve proposal, evidence, approval, execution, outcome, and appeal path.
- **Restore authority:** Align approval rights with the person or role that actually accepts responsibility.
- **Restore contestability:** Create a visible process for appeal, correction, and human explanation.
- **Restore capacity:** Reduce overload, provide better context, improve timing, and remove unnecessary friction.
- **Reduce dependency capture:** Design the support system to teach, scaffold, and return agency rather than permanently replacing the human function.

11. Continuity and Re-Audit

Agency preservation must be tested over time. A system can pass an initial audit and still decay through stale authority, rubber-stamp habits, scope drift, receipt gaps, ignored escalations, or policy-reality divergence.

For that reason, the HAPI Audit Model includes continuity review. Continuity review asks whether the system remained coherent across repeated actions. It does not create a separate governance layer. It strengthens the audit by testing whether the restored conditions stayed restored.

- Did authority remain current?
- Did approvals happen before execution?
- Did receipts remain complete and accessible?
- Did refusals change outcomes?
- Did the system learn from escalations?
- Did support reduce dependency over time?
- Did policy continue to match real operation?

A system that remembers without oversight can drift. A system that never remembers cannot mature. A true system remembers under authority.

12. Applications Across Domains

The HAPI Audit Model is domain-general because agency loss appears across many kinds of systems. The same audit dimensions can be applied differently depending on context.

Agentic AI

In agentic AI, the audit asks whether humans remain meaningfully authoritative over delegated operational agency. It tests whether agents can act outside scope, whether approvals are live, whether execution is bound to permits, and whether receipts preserve accountability.

Workplaces

In workplaces, the audit asks whether employees carry responsibility without authority, whether policies match actual incentives, whether reporting systems protect refusal, and whether performance systems create agency theater.

Healthcare

In healthcare, the audit asks whether patients can understand options, ask questions, refuse, revise care decisions, access records, and remain participants instead of becoming passive cases inside a complex system.

Religious communities

In religious communities, the audit asks whether spiritual authority restores conscience, repentance, truth, and participation, or whether it becomes a false gate that controls belonging, identity, family access, and recognized standing before God.

Government and public systems

In public systems, the audit asks whether citizens can understand decisions, contest outcomes, access reasons, and avoid being reduced to case numbers, risk scores, or administrative categories.

13. Limits and Failure Modes

The HAPI Audit Model has limits. It must not become the very thing it critiques. An agency audit can itself become agency theater if it turns human agency into a shallow score, extracts information from people without returning power, or gives institutions a new language for control without changing consequence boundaries.

Key audit failure modes include:

- **Score capture:** Reducing agency to a number that can be optimized while reality stays unchanged.
- **Audit theater:** Producing a polished report that does not change authority, refusal, memory, or accountability.
- **Surveillance drift:** Using agency audit data to monitor people rather than restore their participation.
- **Consultant capture:** Making the institution dependent on the auditor instead of building internal agency-preserving capacity.
- **Moral overreach:** Claiming authority over domains that belong to conscience, family, faith, or personal responsibility.
- **Overcorrection:** Adding so much process that the audit creates overload and weakens agency.

The audit must therefore remain restoration-oriented. Its legitimacy comes from whether it gives agency back, not whether it produces impressive terminology.

14. Conclusion

The HAPI Audit Model turns the theory of agency into a practical diagnostic tool. It gives organizations, communities, and builders a way to examine whether their systems preserve meaningful participation or merely simulate it.

The model begins with the human. Can the affected person understand, refuse, revise, remember, contest, and remain accountable? If not, governance is incomplete. If the person is visible but powerless, the system has produced agency theater. If the person carries responsibility without authority, the system has split accountability from control. If the system acts before the human can participate, automation has outrun agency.

A HAPI audit does not end with critique. It creates a restoration map. The goal is to rebuild the conditions under which governance becomes real: clarity, authority, refusal, revision, memory, contestability, capacity, and accountability. This is why HAPI treats governance as the product of agency preservation. The audit identifies where agency has been lost, then helps the system rebuild from that core problem outward.

Appendix A: Audit Worksheet

This worksheet can be used as a first-pass diagnostic before a fuller HAPI audit.

6. Who is affected by the system?
7. What decisions or actions create real consequences?
8. Where is the consequence boundary?
9. Who has authority before the boundary?
10. Can the affected human understand the proposal or decision?
11. Can the human refuse before consequence?
12. Can the human revise or escalate?
13. What records preserve what happened and why?
14. Can the human contest or appeal the outcome?
15. Who is held accountable afterward?
16. Does accountability match real control?
17. Does the support system restore capacity or create dependency?
18. Do policies match real behavior?
19. What must change to restore agency?

Appendix B: Finding Template

A HAPI finding should make the agency failure clear enough to repair.

- **Finding title:** Short name for the agency failure.
- **Agency dimension affected:** Clarity, authority, refusal, revision, memory, contestability, capacity, accountability, timing, dependency, or policy-reality coherence.
- **Affected human or group:** Who loses meaningful participation?

- **Consequence boundary:** Where does the system become consequential?
- **Evidence:** What proves the issue?
- **Agency state:** Preserved, weakened, captured, or restored.
- **Severity:** Low, moderate, high, or critical.
- **Restoration requirement:** What condition must be rebuilt?
- **Recommended action:** What practical change should occur?
- **Continuity check:** How should the system verify that the restoration holds over time?

Appendix C: Glossary

Term	Definition
Agency	Meaningful human participation under conditions of capacity, authority, clarity, refusal, memory, and accountability.
Agency loss	The weakening or removal of meaningful participation while the human may remain present in the system.
Agency theater	A condition where participation is displayed but cannot meaningfully affect outcomes.
Consequence boundary	The point where a proposal, judgment, or workflow becomes committed action or real-world effect.
Contestability	The ability to challenge, appeal, correct, or re-open an outcome.
Dependency capture	A support pattern that makes the human less able to act over time.
Policy-reality split	A mismatch between written governance claims and actual operating behavior.
Restoration map	A practical plan for rebuilding the conditions that make agency real.
True gate	A boundary that preserves rightful passage, authority, refusal, memory, and accountability.
False gate	A boundary that captures or controls agency without rightful authority.