

# HAPI IMPLEMENTATION GUIDE

## *From Agency Audit to Restoration, Governance, and Infrastructure*

A Human Agency Preservation Infrastructure Thesis Paper

Author: Michael Bower

Project: Human Agency Preservation Infrastructure (HAPI)

Version: v0.1

Status: Working research manuscript and practical operating guide, not peer reviewed

*Core thesis: HAPI becomes real when agency preservation can be translated into a repeatable process: diagnose agency loss, identify false gates, measure participation, recommend restoration, install governance controls, and verify whether agency improved.*

### **Abstract**

Human Agency Preservation Infrastructure (HAPI) is not complete until it can be practiced. The theory of agency explains what must be preserved; the audit model explains how agency loss can be detected; the business and foundation model explains how the mission can survive institutionally. This implementation guide translates those layers into a repeatable operating method. It proposes a practical sequence for conducting HAPI work: define the system boundary, map the human agency chain, detect agency loss and agency theater, identify false gates and true gates, measure participation quality, produce an agency restoration plan, install governance controls, create evidence, and reassess outcomes over time. The guide treats governance as a product of agency preservation, not a substitute for agency. A HAPI implementation should not merely add policies, dashboards, training modules, or approval checkboxes. It should restore live human authority, refusal, understanding, memory, accountability, and meaningful participation where those capacities have been weakened or removed. This paper provides a field-level operating model for applying HAPI to AI systems, workplaces, healthcare, education, public institutions, religious organizations, families, and other high-pressure systems. Its central claim is practical: if agency can be lost through system design, then agency can also be restored through system redesign.

Keywords: human agency, agency audit, agency restoration, governance infrastructure, false gates, agency theater, participation quality, AI governance, institutional repair, HAPI implementation.

# 1. Purpose of the Implementation Guide

HAPI begins with a simple question: is human agency being preserved, restored, captured, or stripped away? The implementation guide exists to turn that question into a usable process. It gives HAPI a field method. Without a repeatable method, HAPI remains a theory that explains system failure but does not yet repair it.

The purpose of implementation is not to make every person maximally independent. Humans are relational, embodied, limited, and often in need of support. The goal is to distinguish support that strengthens agency from control that replaces agency. HAPI does not reject scaffolding, tools, institutions, or automation. It asks whether those supports help a person or community regain the capacity to understand, choose, refuse, participate, remember, and act responsibly.

*HAPI implementation is the disciplined practice of restoring agency where systems have preserved human presence while removing human participation.*

## 2. The Operating Sequence

A HAPI implementation should follow a clear sequence. The sequence prevents the work from becoming vague moral commentary or ordinary compliance review. Each step produces evidence, decisions, and a concrete restoration path.

1. Define the system boundary: identify the institution, workflow, tool, policy, or relationship being evaluated.
2. Map the agency chain: identify who is meant to understand, choose, refuse, authorize, act, remember, and be accountable.
3. Detect agency loss: find where participation has been replaced by pressure, confusion, automation, dependency, fear, or procedural theater.
4. Identify gates: locate the boundaries where decisions, permissions, consequences, or exclusion occur.
5. Classify gates as true, false, weak, missing, captured, or theatrical.
6. Measure participation quality: evaluate whether human involvement is meaningful or merely present.
7. Write the restoration plan: define what must be rebuilt, removed, slowed down, clarified, returned, or protected.
8. Install governance controls: implement policies, tooling, approvals, receipts, training, escalation, and review mechanisms that preserve agency.
9. Verify outcomes: reassess whether agency actually improved, not merely whether the process became more documented.
10. Create continuity: make sure the system can remember findings, learn safely, and avoid returning to agency loss.

The sequence is intentionally practical. It can be used in a small family system, a workplace workflow, an AI deployment, a hospital process, a school policy, or a public agency. The scale changes, but the core agency question remains stable.

### 3. Phase One: Define the System Boundary

The first implementation failure is trying to audit everything at once. HAPI must begin with a bounded object. The object can be a process, institution, agentic AI workflow, healthcare pathway, educational policy, workplace role, public service, religious discipline process, or family decision pattern.

The boundary should be narrow enough to evaluate and broad enough to include consequence. For example, auditing a chatbot alone may be too narrow if the chatbot triggers real actions through an approval system. Auditing an entire hospital may be too broad if the actual agency loss occurs in patient intake, discharge planning, or insurance navigation.

A good system boundary answers five questions:

- Who is affected?
- What decision or consequence occurs?
- Where does human participation enter the process?
- Where can the process no longer be changed?
- What evidence remains after the process acts?

*A HAPI audit begins where consequence becomes real.*

### 4. Phase Two: Map the Agency Chain

The agency chain is the path from intention to consequence. It identifies where meaning becomes action and who retains authority along the way. In an AI system, the chain may run from human intent to agent proposal, PGDL challenge, action authorization, runtime binding, execution, receipt, and review. In a workplace, it may run from policy to manager instruction, worker task, approval pressure, performance metric, discipline, and appeal. In healthcare, it may run from symptoms to intake, diagnosis, treatment options, insurance approval, consent, follow-up, and recovery support.

A HAPI practitioner should map the agency chain in plain language before recommending any controls. The map should reveal where agency lives, where it transfers, where it is blocked, and where it disappears.

Core agency-chain questions:

- Who initiates the process?
- Who defines the meaning of the situation?
- Who has the power to approve, refuse, revise, delay, or escalate?

- Who is responsible for the consequence?
- Who bears the cost if the system is wrong?
- Who can contest the outcome?
- What proof remains after the action?

The key diagnostic is responsibility without authority. If a person is held responsible for a consequence they cannot understand, influence, refuse, or contest, agency has likely been stripped.

## 5. Phase Three: Detect Agency Loss

Agency loss occurs when a person or group remains present in the system but loses meaningful capacity to participate. They may still click, sign, attend, obey, consent, approve, or perform, but those actions no longer reflect live understanding or authority.

Common signs of agency loss include:

- The person is asked to approve something they cannot reasonably understand.
- The person has no practical ability to refuse.
- The person is pressured to agree quickly before they can deliberate.
- The person is held accountable for a system they cannot influence.
- The process replaces judgment with metrics, scripts, or checkboxes.
- The system creates dependency instead of capacity.
- The human becomes a rubber stamp for a decision already made.
- The institution treats compliance as proof of consent.
- The system removes memory, context, or appeal after consequence.

Agency loss can be soft or severe. Soft agency loss appears as confusion, overload, rushed approval, hidden options, and fragmented memory. Severe agency loss appears as coercion, dependency capture, exclusion, discipline without appeal, irreversible automation, or governance that cannot be contested.

## 6. Phase Four: Detect Agency Theater

Agency theater is the appearance of agency without its function. It occurs when a system preserves the outward symbols of participation while removing the inner conditions that make participation meaningful. It is one of the most important HAPI diagnostics because modern institutions often prefer theater to repair. Theater is cheaper, cleaner, and easier to report.

Examples of agency theater include:

- Human-in-the-loop review where the human cannot change the outcome.
- Consent forms that protect the institution but do not create understanding.
- Training modules that document exposure without building capacity.
- Appeal processes that exist on paper but are practically inaccessible.

- Community feedback sessions where the decision has already been made.
- Safety dashboards that measure activity but not restored judgment.
- AI oversight committees with no authority over deployment speed or scope.
- Religious or institutional discipline processes that call submission voluntary while making refusal socially impossible.

*Agency theater asks whether a human was present. HAPI asks whether human agency was still live.*

## 7. Phase Five: Identify False Gates and True Gates

A gate is a boundary where something is allowed, blocked, escalated, delayed, or transformed. Gates exist in code, institutions, families, religious communities, labor systems, healthcare pathways, and public agencies. HAPI does not assume gates are bad. A true gate can preserve agency by preventing harmful action from outrunning authority. A false gate captures agency by blocking what it has no right to block, permitting what it has no right to permit, or forcing dependence on illegitimate control.

A true gate has five qualities:

- **Rightful authority:** the gate governs only what it has legitimate authority to govern.
- **Reality constraint:** the gate responds to actual risk, truth, context, and consequence.
- **Contestability:** affected people can question, appeal, or correct the gate.
- **Restorative purpose:** the gate aims to preserve or restore agency, not create permanent dependence.
- **Evidence:** the gate leaves proof of why it acted and what conditions applied.

A false gate often looks orderly from the outside. It may have policies, procedures, titles, religious language, compliance language, or technical dashboards. The diagnostic question is not whether the gate looks formal. The question is whether it preserves rightful agency.

## 8. Phase Six: Measure Participation Quality

HAPI measurement should avoid turning agency into a shallow score that creates new compliance theater. Measurement is useful only when it supports judgment. The goal is not to reduce agency to a number. The goal is to make agency loss visible enough to repair.

Participation quality can be evaluated across seven dimensions:

- **Understanding:** does the person know what is happening and why it matters?
- **Choice:** does the person have more than one meaningful path?
- **Refusal:** can the person say no without disproportionate punishment?
- **Revision:** can the person correct, delay, or reshape the process before consequence?
- **Memory:** can the person see what happened, who decided, and what evidence was used?
- **Accountability:** does responsibility match actual authority?

- Restoration: does the process build future capacity or deepen dependence?

A system can pass compliance while failing participation. A system can also be supportive while remaining imperfect. HAPI measurement should separate legal sufficiency from agency health. The central question is whether the system increases or decreases a person's capacity to participate in reality.

## 9. Phase Seven: Write the Agency Restoration Plan

A HAPI audit is incomplete if it only names failure. The output must include a restoration plan. The restoration plan identifies what must be changed so agency can return. This is where HAPI differs from criticism, compliance, and ordinary risk reporting.

Restoration plans should be concrete. They should name the agency loss, identify the affected group, define the failed gate or missing support, propose the repair, assign ownership, and set a reassessment date.

Common restoration interventions include:

- Slow the action boundary so humans can deliberate before consequence.
- Add refusal paths that do not punish legitimate refusal.
- Replace vague approval with informed, scoped, reversible authorization.
- Create receipts so decisions can be remembered and contested.
- Clarify who has authority and who bears responsibility.
- Remove unnecessary dependency on gatekeepers, experts, or automated systems.
- Add support that builds capacity rather than replacing judgment.
- Create human review that can still change the outcome.
- Move from punitive governance to restorative governance where possible.

*The goal of a HAPI restoration plan is not more process. The goal is restored agency.*

## 10. Phase Eight: Install Governance Controls

Governance controls are valid only if they preserve agency. A control that adds paperwork without restoring authority may become theater. A control that blocks all action without building capacity may become dependency capture. A control that produces receipts but cannot be challenged may become surveillance. HAPI controls must be designed as agency-preserving infrastructure.

For agentic AI systems, controls may include:

- Pre-Gate Deliberation Layer review before action proposals become executable.
- Agent Action Gate authorization for consequential actions.
- Runtime Binding to prevent tools from exceeding approved scope.
- Receipts that preserve proposal, objection, approval, permit, execution, and outcome.

- Authority maps defining who can approve which actions under what conditions.
- Escalation paths for sensitive, irreversible, external, or high-impact actions.
- Continuity systems that learn from prior objections and receipts only under human-approved governance updates.

For non-AI institutions, controls may include clearer consent, real appeal paths, role redesign, better information flow, slower irreversible decisions, independent review, training that builds capacity, public reporting, and restored authority for affected people.

## 11. Phase Nine: Verify Outcomes

Verification asks whether agency actually improved. It is not enough to show that the institution adopted the language of agency. It is not enough to show that a dashboard exists. HAPI must measure the difference between declared restoration and lived restoration.

Verification should look for evidence such as:

- People can explain the process more clearly than before.
- Approvals are better informed and less rushed.
- Refusals and escalations are possible without retaliation.
- Irreversible actions require stronger review.
- Responsibility is better matched to authority.
- Receipts make decisions reviewable and contestable.
- Dependency decreases over time because capacity increases.
- Affected people report more clarity, control, and participation.
- The system catches drift before consequence.

Verification must include both structural evidence and human feedback. Structure without lived agency can become bureaucracy. Testimony without structure can become anecdote. HAPI needs both.

## 12. Phase Ten: Build Continuity

A one-time agency audit is useful, but systems drift. Pressure returns. Metrics replace meaning. New tools appear. Leaders change. Incentives shift. HAPI implementation must therefore include continuity. The system should remember what was found, what was restored, what failed, and what needs reassessment.

Continuity should not mean uncontrolled self-modification. In agentic AI systems, receipts and review packets can become governance memory, but policy updates should remain under human authority. In institutions, prior audit findings should become training, policy, and periodic review, not forgotten PDF reports.

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

Continuity turns HAPI from an intervention into infrastructure. It allows systems to mature while preserving human authority over how they mature.

### 13. HAPI Deliverables

A practical HAPI engagement should produce clear deliverables. These deliverables make the work legible to executives, workers, auditors, donors, policymakers, and affected communities.

- System Boundary Brief: what was evaluated and why.
- Agency Chain Map: where intention, authority, action, consequence, and proof flow.
- Agency Loss Findings: where participation is missing, captured, theatrical, or overloaded.
- Gate Classification: true gates, false gates, weak gates, missing gates, and captured gates.
- Participation Quality Review: understanding, choice, refusal, revision, memory, accountability, and restoration.
- Agency Restoration Plan: recommended changes, owners, timelines, and expected agency outcomes.
- Governance Controls Plan: policies, tools, approvals, receipts, training, and escalation paths.
- Verification Report: evidence that agency improved or did not improve.
- Continuity Plan: reassessment schedule and governance memory structure.

These outputs can be adapted to a short public brief, a formal audit report, a certification package, a board memo, or an implementation roadmap.

### 14. Implementation Checklist

The following checklist can guide a first HAPI implementation.

- Define the system boundary and affected population.
- Identify the decision or consequence that makes the system high stakes.
- Map who understands, chooses, refuses, authorizes, acts, remembers, and bears consequence.
- Find points where responsibility exceeds authority.
- Find points where human review cannot change the outcome.
- Find points where support has become dependency.
- Classify true gates, false gates, weak gates, missing gates, and theatrical gates.
- Assess participation quality across understanding, choice, refusal, revision, memory, accountability, and restoration.
- Write concrete restoration recommendations.
- Assign owners, timelines, evidence requirements, and reassessment dates.
- Install or revise governance controls.

- [ ] Verify whether affected people experience more agency after the intervention.

## 15. Application Areas

HAPI can be applied across domains because agency loss is a structural pattern, not a single-sector problem. The language and controls should change by domain, but the diagnostic spine remains stable.

In AI governance, HAPI asks whether agents are acting under rightful authority, whether humans can still intervene before consequence, and whether receipts preserve accountability. In healthcare, HAPI asks whether patients understand, choose, refuse, and navigate care without being crushed by complexity. In education, HAPI asks whether students are building capacity or merely performing compliance. In workplaces, HAPI asks whether employees retain judgment or become rubber stamps for automated systems. In public institutions, HAPI asks whether citizens can contest decisions and participate in systems that affect their lives. In religious institutions, HAPI asks whether shepherding restores agency or false gates capture it.

The point is not to make every domain use the same software. The point is to make every domain answer the same agency question honestly.

## 16. Failure Modes of HAPI Implementation

HAPI must audit itself. Any agency-preservation framework can become a false gate if it becomes too rigid, too proprietary, too performative, or too dependent on its own authority. The implementation model should therefore include self-diagnostics.

- Language capture: institutions adopt HAPI terms without changing power or participation.
- Certification theater: organizations pursue the badge while avoiding restoration.
- Over-gating: HAPI adds so much friction that people bypass it or become dependent on it.
- Consultant capture: the practitioner becomes the permanent gatekeeper instead of restoring local capacity.
- Measurement distortion: agency scores become the goal instead of restored agency.
- Mission drift: revenue incentives soften findings or hide uncomfortable agency loss.
- False neutrality: HAPI avoids naming real harm in order to seem balanced.
- Uncontrolled memory: systems learn from receipts without human-approved governance updates.

*HAPI succeeds only if the implementation restores agency beyond itself.*

## 17. Conclusion

HAPI becomes real when it can move from theory to practice without losing its center. The center is not compliance, software, certification, charity, or branding. The center is preserved and restored human agency.

This implementation guide provides a repeatable path: define the boundary, map the agency chain, detect loss, detect theater, identify gates, measure participation, write a restoration plan, install controls, verify outcomes, and build continuity. The method can be used in AI systems, institutions, public programs, healthcare, education, work, family systems, and spiritual communities because each domain can either preserve agency or strip it away.

The deepest claim is practical: agency loss is not only a philosophical problem. It is a design problem, a governance problem, an institutional problem, and a restoration problem. If systems can be designed to remove agency, they can be redesigned to return it.

*HAPI is the practice of rebuilding systems so people can understand, choose, refuse, participate, remember, and act with restored agency.*

## Appendix A: Field Questions

- Where does consequence become real?
- Who can still change the outcome before that point?
- Who is responsible but lacks authority?
- Who is present but not meaningfully participating?
- What approval is actually a rubber stamp?
- What support is building capacity, and what support is creating dependency?
- What gate blocks without authority or permits without accountability?
- What proof remains after the action?
- What would restored agency look like in this specific system?

## Appendix B: Minimal HAPI Audit Template

A minimal HAPI audit can be written in nine sections:

- System evaluated
- Affected people or roles
- Decision or consequence at stake
- Agency chain map
- Agency loss findings
- Gate classification
- Participation quality assessment
- Restoration plan
- Verification and continuity plan