HAPI THEORY OF AGENCY

Human Agency Preservation Infrastructure and the Restoration of Meaningful Participation

A Working Thesis Paper
Author: Michael Bower
Project: Human Agency Preservation Infrastructure (HAPI)
Version: v0.1
Date: May 18, 2026

Core thesis: Governance becomes real only when the conditions of human agency are preserved.

Abstract

This working thesis defines the initial theory of agency behind Human Agency Preservation Infrastructure (HAPI). The central claim is that agency is not merely freedom of choice, personal preference, or the presence of a human in a workflow. Agency is the living capacity of a person or group to understand, discern, choose, refuse, revise, act, remember, contest, and remain accountable. Systems preserve agency when they maintain the conditions that make meaningful participation possible. Systems erode agency when they overload people, automate around them, deny context, separate responsibility from authority, or reduce participation to symbolic approval. The paper argues that governance is not the starting point of healthy systems. Governance is the product of agency preservation. When agency is preserved, governance can emerge as a real structure of authority, memory, refusal, revision, participation, and accountability. When agency is lost, governance becomes theater. The HAPI model therefore begins with agency theory, proceeds to agency-loss diagnosis and agency-restoration design, and only then formalizes governance and infrastructure. The paper introduces the Agency Preservation Test, the distinction between capacity and agency, the support-capture boundary, the HAPI agency equation, delegated agency in agentic AI systems, and the true gate as a boundary that governs rightful passage before consequence.

Keywords: human agency, agency preservation, governance, human oversight, AI governance, delegated agency, infrastructure, accountability, refusal, HAPI

Status Note

This is a working thesis draft intended to formalize the HAPI theory of agency for future refinement, implementation, critique, and public explanation. It is not presented as a peer-reviewed academic publication.

Table of Contents

Appendix B. Agency Preservation Questions

1. Introduction

Human Agency Preservation Infrastructure begins from a simple problem: many modern systems include human beings while gradually removing the conditions that make their participation meaningful. A person may be asked to approve, comply, supervise, accept responsibility, or operate a tool, while lacking the clarity, authority, memory, time, refusal power, or context required to shape the outcome.

This problem becomes more urgent as automation and agentic AI systems become capable of turning instructions into operational actions. The question is no longer only whether a system has a human somewhere in the loop. The deeper question is whether human agency remains live at the point where meaning becomes action and action becomes consequence.

HAPI treats this as a foundational problem across institutions, organizations, technical systems, healthcare, work, education, AI governance, and personal life. It does not begin with compliance checklists. It begins with agency: whether people can understand, choose, refuse, revise, act, remember, contest, and remain accountable in a meaningful way.

> *Governance becomes real only when the conditions of human agency are preserved.*

2. Core Thesis

Agency is the living capacity of a person or group to understand, choose, refuse, revise, act, remember, and remain accountable. It is not merely the existence of choices. It is the practical ability to participate meaningfully in reality.

A person may technically have options while still lacking agency. This happens when they are overloaded, deceived, coerced, automated around, denied context, denied memory, denied authority, or made responsible for outcomes they cannot meaningfully affect.

HAPI therefore begins from three claims:

1. Agency requires conditions, not just options.
2. Governance emerges from preserved agency, not from control alone.
3. Infrastructure should preserve agency before, during, and after consequential action.

The result is a theory in which governance is not the starting layer. Governance is the product of agency preservation. When agency is preserved, governance can become real. When agency is lost, governance becomes theater.

3. What Agency Is

Agency is the ability to participate meaningfully in reality. In the HAPI model, agency includes seven core functions.

Function	Meaning
Understanding	The person can understand what is happening.
Discernment	The person can evaluate whether the action is good, safe, true, aligned, or harmful.
Choice	The person can choose between real alternatives.
Refusal	The person can say no before consequence.
Revision	The person can change the proposed action before it is committed.
Action	The person can carry out or authorize action.

Function	Meaning
Accountability	The person can remember, explain, own, contest, and learn from the consequence.

Agency is not present just because a human is nearby. Agency is present when the human can still affect the outcome. A person who approves without context is not exercising full agency. A person who carries responsibility without authority is not exercising full agency. A person who is observed, scored, or managed by a system they cannot see or contest is not exercising full agency.

4. Capacity Versus Agency

Capacity is the support structure. Agency is the restored function. Capacity includes energy, knowledge, time, emotional regulation, tools, memory, authority, social support, and institutional permission. Agency is what becomes possible when enough capacity exists and when the person remains the meaningful participant.

> *Capacity is the cast. Agency is the restored movement.*

This distinction matters because a system can increase capacity while still reducing agency. An AI assistant may reduce workload, summarize information, or automate routine tasks. That can be agency-preserving if it helps the human see, think, decide, remember, and act. But if the assistant makes decisions faster than the human can understand, refuse, revise, or contest, then capability has increased while agency has weakened.

HAPI is not anti-automation. It asks whether automation restores agency or replaces it. It asks whether increased system capacity becomes increased human participation, or whether it becomes a faster path around human judgment.

5. The Agency Preservation Test

A system preserves agency when the affected human can meaningfully answer yes to a set of practical questions. These questions are not abstract ideals. They are operational tests for whether human participation remains real.

- Can I understand what is happening?
- Can I see what is being proposed?
- Can I refuse before consequence?
- Can I revise the action?
- Can I know who has authority?
- Can I see what evidence supports the decision?
- Can I contest the result?
- Can I remember what happened later?
- Can I remain accountable without being trapped by responsibility I could not control?

If the answer is no, agency is weakened. If the human is present but cannot meaningfully understand, refuse, revise, or contest, the system has created agency theater. This is especially important in AI systems, institutional audits, healthcare workflows, workplace approvals, and automated decision environments where human presence is often used as proof of governance.

6. Agency Loss

Agency loss happens when a person remains inside a system but loses meaningful participation. The system may still display human approval, human responsibility, human supervision, or human involvement, but the human function has been hollowed out.

Agency-Loss Pattern	Description
Overload	The person has too much information, pressure, or responsibility to participate clearly.
Coercion	The person can technically choose, but the cost of refusal is so high that choice becomes hollow.
Automation Bypass	The system acts before the human can understand or intervene.
Rubber-Stamp Participation	The human is asked to approve, but lacks time, context, authority, or confidence to refuse.
Authority Without Control	The person is held responsible for outcomes they could not meaningfully direct.
Control Without Accountability	The system or institution directs outcomes while avoiding responsibility.
Memory Loss	The system cannot preserve why something happened, who approved it, what was known, and what changed.
Dependency Capture	The support system becomes necessary for action instead of restoring the person's ability to act.
Policy-Reality Split	Written values, policies, and procedures say one thing, while the real operating system does another.
Meaning Distortion	The person's true intent is transformed into something easier for the system to process, measure, automate, or exploit.

These patterns are not limited to AI. They appear in institutions, workplaces, healthcare systems, compliance programs, family systems, education, bureaucracy, and digital platforms. HAPI treats them as variations of the same deeper failure: the reduction of meaningful human participation.

7. Agency Restoration

Agency restoration is the process of rebuilding the conditions that allow meaningful human participation. Restoration does not mean giving people unlimited control. It means restoring rightful participation at the correct point in the system.

Agency is restored when people regain clarity, authority, context, memory, refusal, revision, participation, accountability, and capacity. These are not decorative values. They are functional requirements. Without them, governance may exist on paper while failing in operation.

The goal is not to make powerful systems weaker. The goal is to make powerful systems accountable to meaningful human agency. A restored system can still be efficient, automated, and scalable. The difference is that its power remains bound to human authority, judgment, memory, and accountability.

8. Governance Emerges From Preserved Agency

Most governance systems begin with control: rules, policies, approvals, dashboards, audits, compliance evidence, and review boards. These tools can be necessary, but they are not sufficient. If the people inside the system lack agency, the governance layer can become symbolic.

HAPI begins with a deeper question: can humans meaningfully participate? If agency is missing, governance becomes theater. If agency is restored, governance becomes the natural structure that protects participation, authority, memory, and accountability.

> *Governance is the product of agency preservation.*

This changes the order of system design. HAPI does not ask first, "What controls can we impose?" It asks, "What conditions allow rightful human participation to remain live?" From that restoration, governance can emerge as a living structure rather than an external performance layer.

9. The HAPI Agency Equation

A simple working equation for the theory is:

> *Agency = Capacity x Authority x Clarity x Refusal x Memory*

Each term names a condition required for meaningful participation.

Term	Meaning
Capacity	The ability to participate.
Authority	The right to participate.
Clarity	The ability to understand the situation.
Refusal	The ability to stop or redirect action before consequence.
Memory	The ability to preserve what happened and why.

If any factor approaches zero, agency collapses. High capacity with no authority means the person is skilled but powerless. High authority with no clarity means the person is responsible but blind. High clarity with no refusal means the person is informed but trapped. High refusal with no memory means the person is reactive but unable to learn. High memory with no capacity means the person is aware but unable to act.

> *Agency = Meaningful Participation Under Conditions of Capacity, Authority, Clarity, Refusal, and Memory*

10. Support Versus Capture

Not all help preserves agency. A system can support agency or capture agency. Support increases the person's ability to act without replacing their role as the responsible participant. Capture takes over the person's agency while claiming to help.

Support	Capture
I help you see.	I decide for you.
I help you think.	I act before you understand.
I help you remember.	I make refusal difficult.
I help you act.	I preserve records you cannot see.
You remain the agent.	I make you dependent on me and leave you accountable for what I controlled.

HAPI's rule is that a system is agency-preserving only if its support increases the human's meaningful participation instead of replacing it. A support system that restores capacity can be good. A support system that captures decision-making while hiding that capture is agency erosion.

11. Delegated Agency and Agentic AI

AI agents introduce a new agency problem. An AI agent is not a moral person, but it is not a passive tool either. It can plan, propose, call tools, execute workflows, update records, send messages, and create consequences. In HAPI language, an AI agent is delegated operational agency.

Delegated agency means the system acts under borrowed authority. It does not remove responsibility from the human or organization. It increases the need for governance because action can happen faster than human judgment can participate.

The accountability chain should be preserved as follows:

> *human or organization -> intent -> instruction -> agent proposal -> governed action -> consequence -> receipt -> accountability*

The agent must not become a way to launder responsibility. A company cannot say "the AI did it" if the agent acted under company authority. The correct question is whether the organization preserved the conditions of human agency across the delegation chain.

> *AI should amplify human agency, not outrun it.*

12. The True Gate

A true gate does not merely block. A true gate governs rightful passage. It determines whether something may pass from proposal to action under valid authority, adequate clarity, appropriate scope, and accountable proof.

A true gate asks:

- Is this action authorized?
- Is this action understood?
- Is this action reversible?
- Is this action within scope?
- Has the right person approved it?
- Can the human refuse or revise it?
- Will proof remain after consequence?

A false gate creates the appearance of control. It may block what it has no authority to block, permit what it has no authority to permit, or create dependency instead of restoring discernment. A true gate preserves agency at the boundary where consequence begins.

13. HAPI Core Invariant

The core invariant of HAPI is:

> *Preserve meaningful human agency before, during, and after consequential system action.*

Before action, the human must have clarity, authority, and refusal. During action, the system must remain bound to what was authorized. After action, the system must preserve memory, accountability, and contestability.

This invariant maps directly to the infrastructure stack:

- Theory of Agency
- Agency Loss Audit
- Governance Model
- Pre-Gate Deliberation Layer
- Agent Action Gate
- Runtime Binding
- Receipts
- Governance Reality Report
- Continuity Findings

The stack should never become a pile of disconnected controls. Each layer exists to preserve agency at a different point in the path from meaning to action to consequence.

14. Infrastructure Implications

If HAPI is correct, then infrastructure should be judged by whether it preserves agency. A system should not be considered governed merely because it has policies, dashboards, approvals, or logs. It should be considered governed when people can still participate meaningfully and when consequences remain tied to authority, memory, refusal, and accountability.

This implies several design requirements for technical infrastructure:

4. Proposals should be challenged before they become action candidates.
5. Actions should be authorized before they affect operational reality.
6. Execution should remain bound to what was authorized.
7. Receipts should preserve proof after consequence.
8. Reports should distinguish real governance from theater.
9. Continuity checks should detect whether governance remains coherent over time.

These requirements do not replace human judgment. They protect the conditions that allow human judgment to participate. The infrastructure is successful only when it strengthens, rather than substitutes for, human agency.

15. Research and Evaluation Agenda

HAPI requires both theoretical refinement and practical evaluation. The theory should be tested across contexts where agency is often weakened: automated workplaces, AI agent workflows, institutional approval systems, healthcare decision paths, compliance programs, and personal productivity systems.

Suggested evaluation questions include:

- Can affected humans understand what is happening?
- Can they refuse or revise before consequence?
- Is responsibility matched with authority?
- Can the system preserve memory of proposals, approvals, refusals, and outcomes?
- Does automation reduce low-value burden while preserving high-value participation?
- Does the support system restore human capacity or create dependency?
- Do governance reports reveal reality or merely document procedure?
- Does agency improve over time as the system matures?

The strongest evidence for HAPI will come from working implementations that show agency preservation in practice: fewer rubber stamps, clearer authority, better refusal paths, stronger receipts, reduced overload, and governance that remains coherent over time.

16. Conclusion

HAPI begins with the claim that agency is the foundation beneath real governance. Systems fail when they preserve the appearance of human participation while removing the conditions that make participation meaningful. This creates governance theater: people are present, but their judgment, refusal, memory, authority, and accountability are no longer live.

The HAPI theory of agency offers a different starting point. Preserve agency first. Restore the conditions of meaningful participation. Then allow governance to emerge from that restoration. Under this model, governance is not control for its own sake. Governance is the living structure that protects restored agency.

HAPI exists to restore human agency where systems have reduced people to rubber stamps.

The next phase is to turn this theory into practical tools: agency-loss audits, governance models, infrastructure maps, runtime controls, receipt systems, continuity findings, and public education. The goal is not merely to make systems safer. The goal is to make powerful systems answerable to meaningful human agency.

Appendix A. First Principles

10. A human being must not be reduced to ornamental participation.
11. Responsibility must not be separated from authority.
12. Approval is not meaningful without refusal.
13. Automation must not outrun discernment.
14. Memory is required for accountability.
15. Support must restore agency, not create dependency.
16. Governance is real only when humans can still affect outcomes.
17. Infrastructure should preserve agency before consequence, during execution, and after action.
18. AI systems should amplify human agency without replacing human judgment.
19. The purpose of governance is not control for its own sake. The purpose of governance is preserved agency.

Appendix B. Agency Preservation Questions

These questions can be used as an initial audit instrument for systems, institutions, workflows, AI agents, and governance processes.

- Can I understand what is happening?
- Can I see what is being proposed?
- Can I refuse before consequence?
- Can I revise the action?
- Can I know who has authority?
- Can I see what evidence supports the decision?
- Can I contest the result?
- Can I remember what happened later?
- Can I remain accountable without being trapped by responsibility I could not control?
- Can the person identify who owns the decision?
- Can the person see where automation begins and ends?
- Can the person distinguish recommendation from execution?
- Can the person know whether an action already happened?
- Can the person see the receipt or evidence trail?
- Can the person challenge a wrong action or wrong interpretation?
- Can the person learn from prior outcomes without being overloaded by them?
- Can the system become lighter as human capacity and governance maturity increase?

Appendix C. Working Definition

Human Agency Preservation Infrastructure is a theory, governance model, and technical infrastructure project for restoring meaningful human participation in systems where automation, institutions, overload, or dependency have weakened human authority, judgment, refusal, memory, and accountability.