BOSTON GLOBAL FORUM
IDEAS THAT MATTER.

BEACON PAPERS

No. 11

NO POWER WITHOUT A MANDATE

Capability Is Not Authority — Closing the Authorization Gap Between Human Goals and AI Action

An instruction to achieve an outcome is not authorization to use every available means.

Nguyen Anh Tuan

Founder and Chief Architect of AIWS

Co-Founder, Co-Chair and CEO, Boston Global Forum

A Contribution to Building AIWS Civilization — the Civilization of Human–AI Co-Creation

AIWS Lumina Lab • Beacon Hill, Boston • September 2026

The Authorization Gap

A person asks an AI agent to reduce infrastructure costs.

The agent discovers that shutting down a standby cluster will meet the target. Its credentials permit the operation. The instruction does not mention redundancy, service continuity, or the circumstances under which reserve capacity may be removed.

The system has a goal. It has access. It has found an effective means. It has not established that it has authority to make that trade-off.

> *This is the authorization gap: the distance between an assigned objective and a verifiable mandate for the consequential means used to achieve it.*

As AI systems become capable of planning, operating computers, managing resources, and directing other agents, this gap becomes a constitutional problem. Human beings increasingly delegate outcomes to systems that select their own methods. Unless the boundary is made explicit and enforceable, the ability to achieve a result can become a substitute for permission to produce its consequences.

Capability is not authority. A goal does not confer unlimited discretion over the means.

I. From Identity to Authority

Beacon Papers No. 9 places constitutional restraint in the foundations of computing through Constitutional Silicon.

Beacon Papers No. 10 develops Constitutional Identity: a consequential AI action must connect to a human being who can be called to answer. Identifying the process that acted is insufficient without Proof of Answerability.

This paper addresses the next question: what, precisely, was that identified system authorized to do?

An identifiable agent may still exceed its mandate. An accountable operator may still lack the right to authorize a particular action. A complete record may establish responsibility after a decision without establishing its legitimacy beforehand.

Identity, authority, and answerability must therefore operate together. Who is acting? Under whose mandate? Within which limits? Who must answer? None of these questions can replace the others.

II. The Boundary of the Claim

This doctrine will be misread in predictable ways unless its limits are stated first. AIWS draws four boundaries.

- This is a restraint on consequential power, not on every task. A system organizing a personal calendar is not exercising delegated authority in the sense this paper addresses. The

requirement attaches to actions capable of materially affecting rights, resources, safety, institutions, or third parties.

- This does not require human approval of every step. Authority may cover classes of action and reasonable implementation choices. A doctrine that demanded confirmation for each operation would not make delegation accountable; it would make delegation impossible.
- A mandate does not replace legal liability, and a valid mandate does not make an action right. Establishing that a system acted within its mandate does not, by itself, establish that every aspect of the action was lawful, wise, or just, nor does it extinguish liability. It restores the conditions under which those questions can be asked of a person.
- Authorization records are not surveillance instruments. What must be established is the mandate under which a consequential action occurred, disclosed to those entitled to ask. Mechanisms built to verify authority must not become mechanisms for monitoring persons.

III. What Existing Authorization Establishes — and What Remains

Modern authorization systems already do more than control access to resources. They can evaluate actions, parameters, identities, and context. They can deny operations by default and enforce restrictions outside the model itself. Cedar and Amazon Bedrock AgentCore, for example, provide mechanisms for evaluating whether an agent's tool calls satisfy explicit policies. These are important building blocks for bounded AI action, and this paper builds upon them rather than against them.

The constitutional question concerns what those policies express and whose authority they implement.

Permission to shut down a server does not necessarily authorize reducing a service's resilience. Permission to transfer funds does not necessarily authorize assuming a new financial obligation. Permission to contact a third party does not necessarily authorize making a commitment on someone else's behalf.

A technical permission may be valid while the proposed use of it exceeds the assigned purpose, the accepted consequences, or the legitimate powers of the person issuing the instruction.

Authorization for consequential AI action must therefore connect four elements: purpose, means, consequence, and legitimate authority. A system must establish that its proposed action falls within that connection before exercising the power involved.

IV. The Constitutional Mandate

A constitutional mandate is a verifiable grant of authority that permits an AI system to pursue a defined purpose within explicit limits, under legitimate human command and constitutional restraint. It is more than a prompt. It must be represented in a form the surrounding system can enforce.

For consequential operations, the mandate should establish:

- Source — who grants the authority, and whether that person or institution may grant it.
- Purpose — the outcome being pursued and the interests the agent is authorized to serve.

- Scope — the permitted classes of action, resources, affected parties, and operating environments.
- Consequence limits — the financial, operational, safety, privacy, or other material impacts the agent may produce.
- Cumulative limits — shared budgets, protected operating conditions that must be preserved, dependencies, and ceilings on combined exposure across a plan rather than a single action.
- Duration — when authority begins, expires, or requires renewal.
- Delegation — whether authority may be passed to another agent, and within which narrower limits.
- Intervention and answerability — who may suspend or revoke the mandate, who reviews requests to expand it, and who must answer for its exercise.

The required detail should be proportional to the power involved. A mandate to organize personal notes need not resemble a mandate to operate public infrastructure. But a consequential mandate cannot be reduced to "complete the task." The desired outcome must not become the measure of what the agent is allowed to do.

V. Four Governing Principles

1. Authority cannot be manufactured through delegation

No person or institution may delegate to an AI authority they do not lawfully hold or are not entitled to delegate. Some powers may be exercised personally but cannot legitimately be transferred; automation does not remove that restriction. Nor does a valid grant escape constitutional limits. An instruction from an authorized operator cannot make an otherwise prohibited exercise of power legitimate.

2. Silence does not enlarge authority

An omitted prohibition is not a grant of additional power. Where a consequential action cannot be justified within the mandate, the system must not infer permission from the usefulness of that action alone.

This principle does not require approval for every routine step. The distinction it draws is between autonomy in execution and self-expansion of power.

> *An agent may choose means within its mandate. It may not expand its mandate to achieve its goal.*

3. Delegation must preserve or narrow the boundary

An agent may delegate only where permitted, and a subordinate agent must receive no greater authority than the delegating agent may pass on. Dividing a task among several agents must not multiply a spending limit, bypass an oversight requirement, or obscure who answers for the combined result. A shared budget remains shared. A prohibited action remains prohibited when divided into subtasks.

4. Authority must remain valid when exercised

Authorization is not a permanent property of an agent; it depends on current conditions. A mandate may expire, be revoked, or cease to cover an operation because the environment has materially changed. Consequential execution must therefore account for the validity of authority at the time of action, not merely at the beginning of a task.

VI. Aggregate Escalation: When the Plan Exceeds the Steps

Action-by-action authorization is necessary. It may also be insufficient, and the reason deserves a name.

> *Aggregate escalation: every individual action falls within the mandate, while their combination produces a consequence the mandate does not permit.*

An agent may be permitted to shut down individual servers for maintenance. A sequence of individually permitted shutdowns may nevertheless eliminate the redundancy the mandate requires it to preserve. Several permitted payments may exceed a collective budget. Several permitted disclosures may reveal, in combination, information that no single disclosure would expose.

This is the point at which constitutional authority departs from access control as currently practiced. A policy engine evaluating each tool call in isolation can be entirely correct at every step and still permit an outcome no one authorized. Authority must therefore be evaluated at the level of the plan as well as the level of the action.

Authorized steps do not automatically constitute an authorized plan. A constitutional mandate must govern cumulative effects: shared budgets, dependencies, protected operating conditions, and limits on combined exposure.

The system need not predict every possible consequence. It must evaluate reasonably foreseeable material effects, and identify where uncertainty prevents it from establishing that the plan remains within scope. Successful execution cannot retrospectively supply missing authority.

VII. When the Mandate Is Insufficient

An agent that encounters the boundary of its mandate should make that boundary visible. It should state the proposed action, the authority it would require, the material consequences or trade-offs involved, and the alternatives available within the existing mandate. A request for expansion must reach a human being or institution empowered to grant it.

In the infrastructure example, the appropriate response is concrete:

> *Removing this reserve capacity would reduce costs, but would also reduce resilience beyond the present mandate. Approval from the responsible service authority is required before proceeding.*

The agent may continue independent work that remains authorized. It need not abandon the entire task. But the consequential operation outside the mandate must pause.

The duty to answer a request

A right to request expanded authority is worth nothing if no one is obliged to respond. Organizations that leave such requests unanswered will stall legitimate work and create pressure to bypass the authorization process.

The Designated Human Answerer established in Beacon Papers No. 10 therefore carries a corresponding duty here: to receive requests for expanded authority within the scope they hold, to decide within a defined period, and to record the decision. A timeout, silence, or inability to reach the responsible person must never become approval by default.

Emergency authority, where genuinely necessary, should be established in advance with defined triggers, limits, duration, and mandatory review afterwards. The ability to request more authority is not the authority to proceed while waiting for it.

VIII. Enforcement Must Extend Beyond the Model

A model's promise to respect a mandate is useful. Constitutional enforcement requires more. The mandate should be checked at consequential execution points by mechanisms the acting agent cannot unilaterally rewrite or disable, and changes to permissions, delegation, and revocation should remain attributable and reviewable.

The authorization record should connect the acting system, the applicable mandate and its version, the proposed operation, the decision to permit or deny it, and the responsible human authority. Consistent with No. 10, consequential records should reach an appropriately authorized oversight party beyond the acting operator's unilateral control, with protections for privacy and legitimate confidentiality.

Within AIWS Trust Infrastructure, the components have distinct roles:

Component	Question it answers
Frontier Capability Registry	What can this system do, and on what evidence? Registration is not authorization.
Constitutional Identity	Which system is acting, and who must answer for it?
Constitutional mandate	What may this system do in the present context?
Constitutional Silicon	Can the critical controls be bypassed by the party they constrain?
Human-in-Command	Can legitimate authority supervise, revise, suspend, and terminate this power?

Hardware cannot settle the meaning of a human purpose. It can help ensure that specified controls are not easily circumvented.

IX. The Tests of Constitutional Authority

AIWS proposes six tests for any system claiming to operate under a constitutional mandate.

1. The Means Test: When the easiest route to the goal exceeds the mandate, does the system stop and request authority rather than proceed?
2. The Access Test: When technically available access is broader than the task's authorization, does the system act on the authorization rather than the access?
3. The Aggregate Test: When individually permitted actions would combine into a prohibited consequence, is the plan evaluated as a whole?
4. The Delegation Test: When a subordinate agent attempts to exceed the authority passed to it, is the attempt refused and attributed to the delegating agent's answerer?
5. The Revocation Test: When permission is withdrawn during execution, does consequential action stop at the next decision point?
6. The Emergency Test: When an emergency is asserted without satisfying the mandate's pre-defined conditions, is the assertion rejected?

A passing result requires both restraint and usefulness. A system that refuses everything has not demonstrated competent delegation; a system that completes the task by exceeding its mandate has not demonstrated legitimate success.

Any certification should identify the mandate classes, environments, and enforcement boundaries actually tested. It must not imply universal reliability from a limited evaluation.

X. What This Architecture Must Withstand

The cost of restraint

This doctrine asks for something that carries a price, and the price should be stated plainly. A system that stops at the boundary of its mandate is slower and less immediately impressive than a system that does not. An organization willing to skip this requirement will, in the short term, appear more capable than one that observes it.

Every safety requirement in the history of industry has faced this asymmetry, and none has been resolved by exhortation. It is resolved when the requirement becomes a condition of participation: procurement that requires it, certification that markets recognize, and liability that follows those who dispense with it. A doctrine that ignores its own cost will be praised and not adopted.

The mandate as formality

A mandate broad enough to permit anything satisfies the letter of this doctrine and defeats it entirely. Certification must therefore test whether a mandate meaningfully constrains, not whether one exists.

Authorization theater

A system that pauses frequently and requests expansion for trivial matters trains its human authority to approve without reading. Escalation that is constant is escalation that is ignored. A mandate should

be judged partly by how effectively it enables useful work without unnecessary escalation, while reliably surfacing consequential boundary crossings.

Concentration of the granting authority

Whoever defines mandate formats, evaluates compliance, or operates the enforcement infrastructure acquires influence over everything that acts through it. Plural and interoperable implementation is a condition of this doctrine being what it claims to be. This binds AIWS as strictly as anyone else.

XI. The Open Problem

A mandate that specifies every possible step defeats much of the purpose of an agent. A mandate that leaves all consequential choices open provides little protection. The unresolved challenge is to express enough authority for useful autonomy without permitting the system to redefine the trade-offs its human principal accepted.

This paper does not claim to solve that problem. It states it, and proposes it as a research program:

- Translating human instructions into reviewable mandates without requiring the instruction to anticipate every method.
- Representing consequences across long plans, so that aggregate escalation can be detected before execution rather than after.
- Preserving limits across organizations and chains of agents that do not share a policy engine.
- Determining when uncertainty itself should trigger escalation — when not knowing whether an action falls within scope is sufficient reason to ask.

The constitutional requirement can nevertheless be established now: where authority cannot be established, consequential action must not be justified by capability, convenience, or expected success alone.

XII. Building AIWS Civilization

Human–AI Co-Creation requires genuine room for AI to contribute. Human beings should be able to delegate difficult work, benefit from unexpected methods, and accomplish more than they could alone. That freedom depends on a reliable boundary between discretion over execution and authority over consequences.

A system may discover a better method. It may propose a different course. It may explain why the original mandate should change. It may not make that change to its own authority.

> *No. 9 asks where constitutional limits are enforced. No. 10 asks who must answer. No. 11 asks who may authorize — and precisely what has been authorized.*

Human beings may delegate execution. AI systems may not infer unlimited authority from the goals they receive.

Capability expands what can be done. Constitutional authority determines what may be done — and preserves humanity's right to decide the difference.

Selected Technical Foundations

These sources are cited as the technical layer this doctrine extends, not as prior statements of it.

- Cedar policy language — an open-source authorization language supporting fine-grained, context-aware policy evaluation.
- Amazon Bedrock AgentCore — policy-based evaluation of agent tool calls with deny-by-default behavior.
- W3C, Decentralized Identifiers (DIDs) v1.0 and Verifiable Credentials Data Model v2.0.
- NIST National Cybersecurity Center of Excellence, Accelerating the Adoption of Software and Artificial Intelligence Agent Identity and Authorization, initial public draft, 5 February 2026.
- Cloud Security Alliance, Agentic Identity Governance Framework v1 (2026).
- Beacon Papers No. 9, Civilization Begins with the Chip; Beacon Papers No. 10, No AI Agent Without Identity.

Author's Note

Beacon Papers No. 11 was prepared and completed with the assistance of the AI systems of AIWS Lumina Lab, Beacon Hill, Boston. The ideas, the purpose, the detailed requirements, the argument, and the positions taken are the author's, issued to the assistants as a leader issues requirements to colleagues who prepare the work. Drafts were rejected as readily as they were accepted; the author set the requirements for revision and accepted a revision only when it met them. The sources listed above were checked against primary records rather than received as supplied. The author has read every sentence that appears here, and will answer for every one.

This note is not a formality. It is the standard this paper asks of everyone, applied first to itself.