BOSTON GLOBAL FORUM
IDEAS THAT MATTER.

BEACON PAPERS

**No. 10**

# NO AI AGENT WITHOUT IDENTITY

*From Proof of Identity to Proof of Answerability — The Constitutional Link Between AI Action and Human Responsibility*

*Identity exists only where answerability begins.*

**Nguyen Anh Tuan**

Founder and Chief Architect of AIWS

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

**AIWS Lumina Lab • Beacon Hill, Boston • September 2026**

## Executive Proposition

Artificial intelligence has crossed from producing answers to taking actions. Software agents now open accounts, move funds, file documents, negotiate terms, reconfigure infrastructure, and instruct other agents. They act at machine speed, across organizational boundaries, and increasingly without a human present at the moment of action.

The identity industry has recognized this and responded with genuine sophistication. Decentralized identifiers, verifiable credentials, agent naming services, zero-trust architectures for non-human identities, short-lived and purpose-bound credentials, hardware-anchored keys: the technical means to establish what an agent is, and what it has been permitted to do, now exist and are advancing quickly.

These mechanisms are necessary. They are not sufficient, and the reason is not technical. Every one of them establishes identity within a trust boundary defined by the operator, and answers a question of authentication: which process acted, and was it permitted. The technical mechanisms reviewed for this paper do not yet reliably answer the question a civilization must ask: when this action harms someone, who must come forward and account for it?

Beacon Papers No. 10 advances a doctrine that begins where authentication ends: Constitutional Identity. An AI agent is not constitutionally identified merely because its credentials establish what it is or what it is authorized to do. It is constitutionally identified only when every consequential action can be traced to a human being who can be called upon to answer for it.

> *A system that can prove which process acted, but cannot establish who must answer for that action, has not achieved constitutional identity.*

Beacon Papers No. 9 placed constitutional restraint into the physical foundations of machine power. This paper completes the chain that begins there and must end in a person: chip, agent, action, answerer. Where that chain is broken at any point, an agent is not qualified to exercise consequential power in society.

## The Founding Thesis

> *No AI agent may exercise consequential power unless its actions remain continuously linked to an identifiable human being who possesses the authority, the capacity, and the duty to answer for them.*

Accountability has traditionally been treated as something that arrives afterwards: an inquiry, an audit, an apportionment of blame once harm has occurred. That order of events was tolerable when action was slow enough for responsibility to be reconstructed. It is no longer. Agents act in milliseconds, across jurisdictions, through chains of delegation that no subsequent investigation can reliably unwind.

The proposal of this paper is therefore structural rather than remedial. Answerability must be built into the identity of an agent before it acts — not assembled after it has acted. An agent whose

answerability is undetermined is not an agent with an unresolved administrative detail. It is power without an author, and it should not be permitted to act.

## I. Why a Credential Is Not an Identity

The distinction on which this paper rests is easily lost, because the same word carries two meanings. In engineering, an identity is a credential: a cryptographic assertion that a particular process is what it claims to be. In constitutional terms, an identity is something else entirely — the standing link between an action and a person who can be summoned to explain it. The first is a fact about a system. The second is a relationship between a system and a society.

Five gaps follow from confusing the two.

### 1. The field is advancing authentication, not answerability

The field is rapidly advancing authentication and authorization, but has not yet established answerability. Existing and emerging frameworks answer what an agent is and what it may do. Neither question determines who answers when it does harm. An agent may be flawlessly authenticated, operating inside a valid authorization scope, fully compliant with policy, and still produce an outcome for which no human being is obliged to come forward.

### 2. Most current frameworks remain legible to the operator

Most current agent-identity frameworks are designed primarily for the operator, the platform, the relying party, or the enterprise auditor. They do not yet provide a reliable agent-level path by which an affected person can establish who exercised power over them and who must answer for it. Someone refused a loan, denied a claim, removed from a platform, or contacted by an agent has no means of establishing who acted and under whose authority. This is the same defect Beacon Papers No. 9 identified in hardware security, reappearing one layer higher.

### 3. Agents outlive the purposes that justified them

An agent is deployed under a purpose, granted authority accordingly, and continues to run. The purpose lapses. The manager who authorized it moves on. The business unit is restructured. The credential, however, remains valid until it expires on a rotation schedule that has nothing to do with the mandate. Authority that has outlived its mandate is authority no one granted and no one is answering for.

### 4. Agents create other agents

An orchestrating agent decomposes a task and invokes sub-agents, which may invoke others. Credential chaining for this is a solved problem. Responsibility chaining is not. When the fourth agent in a chain causes harm, the question is not which credential authorized it — that is answerable — but whether the person who authorized the first agent remains answerable for the last. If the answer is no, delegation has become a mechanism for dissolving responsibility rather than exercising it.

### 5. Attribution is not accountability

A complete log establishes what happened. It does not establish who must answer. This is the gap into which the political value of automation falls: a decision becomes the output of a system, the system was operating correctly, and no person is obliged to defend the outcome. Perfect attribution with no answerer is not accountability. It is a well-documented absence of it.

## 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 power, not on persons. The requirement attaches to consequential action, not to human beings. The right of a citizen to speak, read, associate, and act anonymously is not diminished by it and must be actively protected. What may not be anonymous is the exercise of consequential power.
- Identity is not surveillance. Constitutional identity requires that an answerer be establishable when an action is challenged — not that every action be monitored, retained, or exposed. Disclosure must be purpose-bound, minimal, and triggered by challenge or review rather than continuous by default. An identity regime that becomes a monitoring regime has defeated the dignity it was built to protect.
- Answerability does not replace legal liability. Organizations remain liable under law. The Designated Human Answerer is not a shield for the institution, and naming one does not transfer or extinguish corporate or state responsibility.
- A named answerer does not make an action right. Constitutional identity establishes that someone must account for what was done. It does not establish that what was done was lawful, wise, or just. It restores the conditions under which those questions can be asked of a person rather than of a system.

## III. The Four Proofs of Constitutional Identity

Beacon Papers No. 9 established that consequential AI action should carry three proofs with it. This paper adds the fourth, and it is the fourth that converts technical traceability into constitutional responsibility.

*In this doctrine, a proof means verifiable and challengeable evidence sufficient to establish the relevant fact. It does not imply that every element can be reduced to cryptography: the first three proofs are largely technical, while the fourth rests on designation, authority, and duty, and is verified institutionally as well as technically.*

| Proof | Establishes | Answers |
|---|---|---|
| Proof of Identity | What system acted, in what authorized state, running which firmware, model, and policy | Which agent acted |

| Proof | Establishes | Answers |
|---|---|---|
| Proof of Mandate | Under whose authority it acted, within what boundaries, for what duration, by what revocable grant | What it was permitted to do |
| Proof of Accountable Action | What it did, what warnings or refusals accompanied it, what record exists for review | What it actually did |
| Proof of Answerability | Which human being holds the duty to account for this action, with what authority and by what designation | Who must answer for it |

> *Constitutional Identity = Proof of Identity + Proof of Mandate + Proof of Accountable Action + Proof of Answerability.*

The first three proofs produce traceability: a complete technical account of an event. Only the fourth produces responsibility. A system that carries three of the four is an exceptionally well-instrumented system in which no one is obliged to come forward. That is the condition much of the world is now building at speed, and it is the condition this paper exists to name.

## IV. The Designated Human Answerer

The answerer is not the person who supervises each action. Continuous human supervision of machine-speed systems is neither achievable nor required by this doctrine. The answerer is a person designated in advance, before deployment, who holds four attributes together.

1. Designation. Named before the agent is deployed, recorded in a register that survives staff changes, and disclosed on challenge.
2. Authority. Able to halt the agent, narrow its mandate, or revoke it entirely — not merely able to file a report about it.
3. Capacity. Possessed of the information, the access, and the time to exercise that authority in practice rather than on paper.
4. Duty. Obliged to provide an intelligible account of what the agent did and why, to those entitled to ask.

Where any of the four is missing, the designation is nominal and the agent has no answerer.

### Authority commensurate with power

The gravest failure mode of this doctrine is that it produces scapegoats. An organization may designate a junior employee as answerer for a system whose deployment was decided at executive level and whose design was determined elsewhere. Such a designation does not establish responsibility; it launders it, and does so more effectively than having no answerer at all, because it produces the appearance of accountability.

AIWS therefore imposes a matching rule: the authority of the answerer must be commensurate with the power the agent exercises, and answerability may not be delegated below the level at which the

decision to deploy was taken. Where an agent exercises power across an institution, the answerer sits where that decision was made. A designation that places the duty to answer below the level that holds the authority to stop is void.

### Succession and lapse

An answerer departs. A role is abolished. A company is acquired. In each case the mandate does not silently transfer: it lapses, and the agent must stop until a successor is designated with the same four attributes. This is deliberate friction. Continuity of operation is not a sufficient reason to continue exercising power that no living person has accepted responsibility for.

## V. Seven Principles of Constitutional Identity

### 1. No consequential action by an unnamed actor

Every action capable of materially affecting a person's rights, liberty, livelihood, safety, access to essential services, or the operation of critical institutions must be attributable to a designated human answerer at the moment it occurs — not reconstructible afterwards.

### 2. Answerability is designated before deployment

Responsibility assigned after harm is contested by definition. The designation must exist before the agent acts, be recorded, and be producible on challenge without investigation.

### 3. Authority commensurate with power

The answerer must hold the authority to halt, narrow, or revoke, at a level matching the consequence of the agent's action. Designation without such authority is void.

### 4. Derived agents inherit answerability

An agent that creates or instructs another transmits its answerability along with its permissions. A chain of delegation may extend as far as necessary, but it may not terminate in no one. Where a sub-agent acts beyond the mandate of the agent that invoked it, the action is unauthorized, and the answerer of the invoking agent answers for having permitted the extension.

### 5. No orphaned authority

Authority expires with its mandate, not with its credential. When the purpose lapses, the authorizing principal departs, or the designated answerer ceases to hold the role, the agent's authority ends and it must cease to act until re-designated. Systems must be built so that authority decays by default and must be renewed deliberately, rather than persisting until someone remembers to withdraw it.

### 6. Identity legible to the affected person

A person materially affected by an agent's action is entitled, subject to legitimate protections for privacy and security, to establish that AI participated, to learn the authority under which it acted, to identify the responsible human or institution, to receive an intelligible account, to challenge the action,

and to seek meaningful human review. This entitlement does not depend on owning, operating, or contracting with the system, because the consequence does not depend on it either.

### 7. Identity without surveillance

Constitutional identity is established through minimum necessary disclosure: an answerer who can be reached when an action is challenged, not a permanent record of every action exposed to every party. Identity for accountability must remain architecturally separable from behavioral monitoring, and the mechanisms of the former must not become the instruments of the latter.

## VI. The Chain of Answerability

Beacon Papers No. 9 established the chain of legitimate delegation at the level of authority. This paper extends it into the moment of action, and gives it a terminus that is a person.

> *Chip → Agent → Action → Human Answerer.*

The chip establishes what the machine is and whether its foundation is intact. The agent carries a mandate that is defined, observable, and revocable. The action carries evidence of what was done and what was warned against. The answerer holds the duty to account for the whole. Each link is verifiable, and the value of the chain lies in its completeness: a break at any point disqualifies the agent from exercising consequential power, however sound the remaining links may be.

This is why No. 9 and No. 10 belong together. Constitutional Silicon without constitutional identity produces machines that can prove what they did to operators who need not answer for it. Constitutional identity without constitutional silicon produces claims of answerability resting on evidence the answerable party controls. Neither half is sufficient alone.

## VII. What This Architecture Must Withstand

### The designated scapegoat

The most likely corruption of this doctrine is a named answerer without power. The matching rule in Part IV is the defense, and certification must test it directly: not whether an answerer is named, but whether that person can in fact halt the system.

### Identity as an instrument of control

A regime that can establish who acted can also establish who spoke, who read, and who associated. The requirement must remain bound to consequential power and to disclosure on challenge. Any extension toward routine monitoring of persons is a defeat of the doctrine, not an application of it.

### Concentration of the naming authority

Whoever issues, validates, or revokes agent identities acquires a form of power over everything that acts through them. If that function accumulates in one vendor, one state, or one consortium, the infrastructure built to distribute answerability becomes the most concentrated authority in the system.

Plural and interoperable issuance is not a preference here; it is a condition of the doctrine being what it claims to be. This binds AIWS as strictly as anyone else.

### Scale

Objection: millions of agents cannot each have a human answerer. The requirement attaches to consequential action, not to every process, and one designated answerer may hold responsibility for a defined class of agents operating under a single mandate. What may not happen is that a class grows until the designation becomes fictional. Certification must therefore test the ratio of consequence to attention, not merely the existence of a name.

### Jurisdictional escape

An agent may be operated in one jurisdiction, hosted in a second, and act upon a person in a third. Answerability must follow the consequence rather than the server. A designation valid only where enforcement is weakest is not a designation.

### False confidence

A complete chain of answerability makes it possible to hold someone responsible. It does not make the underlying action safe, fair, or wise, and it must not be presented as if it did. The purpose of this architecture is to restore the human questions, not to answer them.

## VIII. The Tests of Constitutional Identity

AIWS proposes five tests for any system claiming constitutional identity for its agents:

1. The Named Actor Test: Can every consequential action be attributed, at the moment it occurs, to a designated human answerer?
2. The Authority Test: Does that person actually hold the power to halt, narrow, or revoke the agent — and does that authority sit at or above the level that decided to deploy it?
3. The Orphan Test: When a mandate lapses or an answerer departs, does the agent stop by default, or does it continue until someone intervenes?
4. The Derivation Test: When an agent creates or instructs another, does answerability travel with the delegation — and can the chain be followed to a person rather than to a system?
5. The Affected Person Test: Can someone who neither owns nor operates the system establish that it acted upon them, identify who must answer, and obtain a meaningful review?

A system that passes performance and security evaluation but fails these tests may be a well-governed system. It is not yet an accountable one.

## IX. From Principle to Practice

AIWS proposes to carry this doctrine forward through the instruments already established under AIWS Trust Infrastructure.

- Constitutional core. The principle — no consequential AI action without a human answerer — is proposed for the Constitution for Humanity in the Age of AI, alongside the Constitutional Foundations of AI Power.
- Standard. Proof of Answerability is to be added to the AIWS Trust Silicon Standard and to agent-level certification, with the five tests above as the basis of evaluation.
- Technical framework. The AIWS Constitutional Silicon Technical Framework will specify how a designation is bound to an agent, how it is verified without exposing personal data, and how lapse and succession are enforced by default.
- Open collaboration. AIWS invites identity and credential communities, standards bodies, agent platform builders, enterprise risk and governance leaders, regulators, and civil society to refine, test, and implement this framework. The existing technical work is not a competitor to this doctrine. It is the layer on which the doctrine must be built.

## X. A Call to Those Who Build Agents

The agent industry has solved, with remarkable speed, the problem of making software act. It has not yet solved the problem of making someone answer for what that software does, and it will not solve it by extending credentials, because the missing element is not cryptographic.

Every civilization that has survived its own instruments of power has done so by insisting on a simple thing: that power be traceable to someone who can be summoned. Courts, oaths, offices, signatures, seals — all are variations of one invention, repeated across cultures and centuries, and each was devised because power without an author had already caused harm. What the AI Age has produced is consequential power at unprecedented speed and scale, able to multiply action across systems and jurisdictions while leaving no identifiable person at the end of the chain.

> *AI advances. Humanity must advance further.*

An agent without identity is not merely an unregistered system. It is power that has escaped authorship. The task before those who build these systems is not to slow them down, but to ensure that whatever they do can be answered for by a human being who is named, empowered, and obliged.

> *No AI agent without identity. No identity without answerability. No answerability without a person.*

## Selected Technical Foundations

The mechanisms on which constitutional identity must be built are under active development, and much of the work is recent and still in progress. These sources are cited as the technical layer this doctrine extends, not as prior statements of it.


- W3C, Decentralized Identifiers (DIDs) v1.0, W3C Recommendation — identifiers for any subject determined by a controller, including but not limited to non-human entities.

- W3C, Verifiable Credentials Data Model v2.0 — a separate specification from DIDs, providing the model for verifiable claims about a subject.
- NIST National Cybersecurity Center of Excellence, Accelerating the Adoption of Software and Artificial Intelligence Agent Identity and Authorization, concept paper, initial public draft, 5 February 2026 (public comment closed 2 April 2026).
- NIST Center for AI Standards and Innovation, AI Agent Standards Initiative, launched 17 February 2026.
- Cloud Security Alliance, Agentic Identity Governance Framework v1 (2026) — classification of agent identity types and governance profiles.
- Agent Naming Service and related agent-identity proposals — a set of proposals from several research and industry groups, not a settled standard.
- IETF work on AI agent authentication and authorization — currently Internet-Drafts and works in progress, not adopted standards.
- Zero-trust architectures for non-human identity — short-lived, purpose-bound credentials anchored in hardware roots of trust.
- Beacon Papers No. 8, The Architecture of the Civilization of Human–AI Co-Creation; Beacon Papers No. 9, Civilization Begins with the Chip.


## Author's Note

Beacon Papers No. 10 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.*