PUBLIC MANIFESTO

The Accountability Invariant

On the boundaries that must not move

Version 1.0 · December 29, 2025 · CC BY 4.0

This document establishes a non-negotiable boundary for artificial intelligence systems that affect human lives.

It is not a product specification. It is not a marketing position. It is not a guideline that admits exceptions.

It is an invariant: a property that must hold regardless of implementation, regardless of capability, regardless of commercial pressure.

This boundary is stated once and must never weaken:

THE CORE INVARIANT

AI must never be the last responsible actor.

AI may inform decisions.
Responsibility must always terminate at a human or an institution.

I. What This Means

When a system powered by artificial intelligence produces an outcome (a recommendation, a classification, a prediction, a decision), someone must be accountable for that outcome.

That someone cannot be the software.

Software has no legal standing. Software cannot be sued, fired, sanctioned, or held to account. Software cannot explain itself under oath. Software cannot feel the weight of having caused harm.

Therefore, software cannot be responsible.

> RESPONSIBILITY
>
> The obligation to answer for outcomes, accept consequences, and provide remedy. Responsibility requires agency, legal standing, and the capacity to be held accountable.

Machines possess none of these. They execute. They do not decide. They produce outputs according to mathematical functions derived from data. The appearance of decision is an artifact of complexity, not evidence of agency.

Any system architecture that obscures this fact (that allows the chain of responsibility to terminate at "the AI") is not merely poorly designed. It is a failure of governance that creates unaccountable power.

II. What This Is Not

This invariant is not:

- **A guideline.** Guidelines admit exceptions. This does not.
- **An aspiration.** Aspirations describe hoped-for states. This describes a constraint that must hold now.
- **A best practice.** Best practices are recommendations. This is a requirement.
- **An opinion.** This is a statement about the structure of accountability, not a preference.

The invariant does not claim that AI is dangerous. It does not claim that AI should be restricted. It claims only that AI cannot be responsible: and therefore, systems that use AI must preserve human responsibility.

This is not written against artificial intelligence.

It is written for humanity.

III. Why This Matters

When responsibility becomes unclear, three failure modes emerge:

Automation Bias

Humans defer to machine outputs because machines appear authoritative. The recommendation becomes the decision. The human reviewer becomes a rubber stamp. The system has, in effect, decided; but no one has accepted responsibility for that decision.

Blame Displacement

When outcomes are poor, institutions point to "the algorithm." The algorithm cannot respond. It cannot be sanctioned. It cannot provide remedy. The harmed party has no recourse. The institution has created a shield of plausible deniability using software.

Silent Escalation Failure

High-stakes decisions pass through automated systems without triggering review. No human sees the edge case. No human overrides the error. By the time harm occurs, no one remembers that a human was supposed to be in the loop.

THE COMMON THREAD

In each case, the failure is not technical. The failure is that responsibility was never clearly assigned: or was assigned to a system incapable of bearing it.

Ambiguity is the real failure mode.

IV. The Structure of Accountability

Accountable systems require explicit answers to four questions:

1. WHO RECOMMENDS?

The AI system. It provides analysis, patterns, predictions, and suggestions. It has no authority.

2. WHO DECIDES?

A human. The human evaluates the recommendation, applies judgment, and makes the final determination.

3. WHO ENFORCES?

An institution. Policy defines the rules. The institution ensures compliance and provides remedy when rules are violated.

4. WHO IS ACCOUNTABLE?

The human and the institution. Never the software. The chain of responsibility must terminate at entities with legal standing.

These roles must be explicit in every system that affects human welfare. They must be documented. They must be auditable. They must not collapse into each other.

FORMAL REQUIREMENT

For any AI-assisted outcome O:

```
∃ human H or institution I such that: accountable(O)
= H ∨ accountable(O) = I There exists no valid state
where: accountable(O) = AI
```

V. Why Mathematical Governance

The language of AI governance is often imprecise:

- "Human-in-the-loop" - but with what authority? At what point? With what training?
- "Responsible AI" - responsible to whom? Measured how?
- "Ethical AI" - whose ethics? Enforced by what mechanism?
- "Trustworthy AI" - trust is an outcome, not a feature.

These phrases describe intentions. They do not describe constraints.

Governance that relies on intention alone fails when intentions conflict with incentives. When the cost of human review is high, humans will be removed from the loop. When the benefit of faster decisions is clear, authority will migrate to automation. Good intentions provide no structural resistance to these pressures.

Mathematical governance is different. It specifies:

- **Conditions** under which escalation must occur: not "should," but "must."
- **Thresholds** that trigger review: measured, not felt.
- **Mappings** from every output to an accountable actor: complete, not aspirational.
- **Auditable logs** that prove compliance: verifiable, not claimed.

THE STANDARD

If a governance rule cannot be expressed formally, it cannot be enforced reliably. If it cannot be enforced reliably, it will fail under pressure. Ethics without enforcement is theater.

VI. On Human Stewardship

Humans remain accountable even when machines are correct.

This may seem counterintuitive. If the AI's recommendation was right, why should the human bear responsibility? The answer is structural: accountability is not about blame for errors. It is about the legitimacy of decisions that affect lives.

A correct recommendation that bypasses human judgment is still illegitimate if the domain requires human authority. A physician who follows an AI suggestion without independent evaluation has not practiced medicine; they have executed an instruction. A judge who defers to a risk score without scrutiny has not rendered judgment; they have automated sentencing.

The requirement for human authority is not about catching AI errors. It is about preserving the meaning of human decision-making in domains where that meaning matters.

> STEWARDSHIP
>
> The obligation to maintain responsibility even when delegation is efficient. Stewardship recognizes that some decisions must remain human not because humans are better, but because the decision itself requires human agency to be legitimate.

Authority must never be automated by default. It may be informed by automation. It may be accelerated by automation. But the moment of decision (the acceptance of responsibility for the outcome) must remain with a human or an institution capable of bearing it.

VII. On Language

Language shapes accountability. Imprecise language enables evasion.

NEVER SAY

- "The AI decided..."
- "The system approved..."
- "The algorithm determined..."

These phrases attribute agency to software. Software does not decide, approve, or determine. Software executes functions.

ALWAYS SAY

- "The AI recommended..."
- "The reviewer approved..."
- "The institution determined..."

These phrases preserve the distinction between computation and authority. They name the accountable party.

This is not pedantry. Language that attributes decisions to AI creates the conceptual space for blame displacement. Once we accept "the AI decided," we have already lost the ability to ask who is responsible.

Precision in language is the first defense against erosion of accountability.

VIII. What This Requires

Systems that honor this invariant must implement:

Explicit Role Separation

Every AI-assisted process must document who recommends, who decides, who enforces, and who is accountable. These roles must not collapse. The person who reviews AI output must have the authority and obligation to override it.

Deterministic Escalation

Conditions that require human review must be specified in advance, measured automatically, and enforced without exception. "High confidence" is not an excuse to skip review if the policy requires it. Escalation triggers must be mathematical, not discretionary.

Auditable Responsibility

Every outcome must map to an accountable actor. This mapping must be logged, preserved, and available for review. When something goes wrong, it must be possible to answer: who was responsible for this decision?

Immutable Records

Audit trails cannot be modified after the fact. The chain of responsibility must be reconstructible from logs alone. If the system cannot prove who was accountable, accountability did not exist.

IX. On Trust

This document does not promise trust.

Trust is not a property that can be declared. It is not achieved by stating "our AI is trustworthy." It is not conferred by certifications or compliance badges.

Trust emerges (or fails to emerge) from the consistent behavior of systems over time, observed by people who have reason to scrutinize them.

What this document promises instead:

- **Transparency** about what the system does and does not do.
- **Structure** that preserves human authority.
- **Mechanisms** that make accountability enforceable.
- **Records** that allow verification.

Whether trust follows is not for us to claim. It is for others to decide, based on evidence.

X. The Final Test

Any system, any organization, any document that claims to govern AI responsibly must pass one test:

THE QUESTION

"Who is accountable when this system fails?"

The answer must be a human or an institution.

The answer must never be:

- "The AI."
- "The algorithm."
- "The model."
- "Nobody. It was automated."

If any architecture, any policy, any process permits these answers, it has failed. If any sentence in any governance document weakens the guarantee that a human or institution will be accountable, that sentence must be removed.

This is the invariant. It does not bend.

This manifesto is offered freely. It may be copied, adapted, cited, and built upon. Its value lies not in ownership but in adoption.

The boundary it describes is not new. It is as old as the concept of responsibility itself. What is new is the pressure that AI systems place on this boundary: the temptation to automate not just tasks but accountability, to create systems that act without anyone to answer for them.

That temptation must be resisted.

Not because AI is dangerous. Not because machines cannot be trusted. But because accountability is the foundation of legitimate authority, and authority that cannot be held accountable is authority that should not exist.

This document is written for engineers building systems, for regulators overseeing them, for executives deploying them, for citizens affected by them, and for historians who will judge whether we preserved what mattered.

The invariant is simple. The obligation is permanent.

> **AI must never be the last responsible actor.**

ISSUED BY

TeraSystemsAI

December 29, 2025

The Accountability Invariant

Version 1.0 · December 29, 2025