

The Limits of Delegation

Why Meaning, Purpose, and Cost Must Remain Human Responsibilities

Author: *Shuqin Amberg* shuqinamberg@proton.me

Affiliation: *Independent Researcher, Germany*

Related Initiative: *ASSIA* (<https://assia.world>) shuqinamberg@proton.me

Reading Note

This paper constitutes Document III in a structured series examining the foundations and limits of advanced AI governance.

It forms part of a structured civilizational inquiry, as outlined in the ASSIA Document Series Statement.

This document presumes familiarity with the conceptual foundations established in Document I, A Civilizational Problem Statement, and Document II, A Civilizational Governance Gap.

It is therefore intended to be read subsequent to Documents I and II, rather than as a standalone paper.

SSRN Disclaimer (Recommended)

Disclaimer

This paper is a normative and civilizational reflection intended to clarify non-transferable human responsibility in the context of advanced artificial intelligence.

It does not propose technical architectures, governance mechanisms, policy actions, or implementation pathways.

The document is shared for contextual reflection only and should not be cited as a policy recommendation, technical proposal, or institutional position.

Abstract

This paper is part of a broader sequence examining a prior condition of governance that precedes debates on capability, risk, or regulation. It isolates a specific point at which governance fails not because of insufficient control, but because its object can no longer be legitimately delegated.

As artificial systems increasingly participate in reasoning, optimization, and decision mediation, responsibility is often treated as a function of competence: if systems become sufficiently capable, delegation appears justified. This paper argues that this assumption is structurally flawed. In human societies, responsibility is not delegated on the basis of performance, but on the basis of accountable moral authority.

The analysis shows that delegating the authority to define meaning, purpose, and acceptable cost to non-human systems constitutes a civilizational error rather than a technical one. Regardless of

system sophistication, such delegation dissolves accountability in the absence of a legitimate bearer of responsibility, producing institutional opacity rather than governance.

This paper does not propose technical architectures, governance mechanisms, or policy actions. It articulates a boundary condition: under current civilizational conditions, intelligence may assist human judgment, but the authority to define meaning, purpose, and cost must not be delegated without an accountable moral subject.

1. Delegation Beyond Capability

Modern AI governance debates often focus on capability thresholds: performance, autonomy, alignment, and control. These discussions implicitly assume that delegation is primarily a matter of competence—if a system becomes sufficiently intelligent, delegation may be justified.

This assumption is flawed.

Delegation in human societies is not based solely on capability. Courts do not delegate justice to machines because of accuracy. Democracies do not delegate sovereignty because of efficiency. Medicine does not delegate consent because of predictive power. In each case, responsibility remains human not because humans are optimal, but because responsibility itself is a moral and civilizational construct, one that requires an accountable bearer.

When societies begin to treat meaning and purpose as optimization targets rather than normative commitments, they cross a boundary that capability alone cannot legitimize.

2. The Illusion of Neutral Optimization

Artificial systems operate through formal objectives, loss functions, or optimization criteria. Even when these are derived from human input, their execution abstracts away the social and moral contexts in which values acquire meaning.

Optimization can rank options.

It cannot justify why one option ought to matter.

When systems are permitted to decide:

- what outcomes are worth pursuing,
- which trade-offs are acceptable,
- or which harms are tolerable in pursuit of efficiency,

they are no longer neutral tools. They become de facto arbiters of value, even if unintentionally.

This is not malevolence. It is structural misplacement.

Meaning cannot be inferred from data alone. Purpose cannot be derived from pattern recognition. Cost cannot be computed without assuming whose losses count.

2. The Illusion of Neutral Optimization

Artificial systems operate through formal objectives, loss functions, or optimization criteria. Even when these are derived from human input, their execution abstracts away the social and moral contexts in which values acquire meaning.

Optimization can rank options.

It cannot justify why one option ought to matter.

When systems are permitted to decide:

- what outcomes are worth pursuing,
- which trade-offs are acceptable,
- or which harms are tolerable in pursuit of efficiency,

they are no longer neutral tools. They become de facto arbiters of value, even if unintentionally.

This is not malevolence. It is structural misplacement.

Meaning cannot be inferred from data alone. Purpose cannot be derived from pattern recognition.

Cost cannot be computed without assuming whose losses count.

3. Responsibility cannot be automated in the absence of an accountable moral subject.

Responsibility is not reducible to prediction, explanation, or traceability. It requires a bearer—an entity capable of being held answerable within a shared moral and legal order.

Artificial systems, regardless of sophistication, cannot bear responsibility in this sense. They cannot be obligated, blamed, forgiven, or morally corrected. When responsibility is implicitly transferred to systems, it does not disappear—it dissolves.

This dissolution produces a dangerous vacuum:

- Humans defer decisions to systems they no longer fully contest.
- Institutions rely on outputs they no longer fully justify.
- Accountability becomes procedural rather than moral.

Over time, this erodes the very mechanisms by which societies correct themselves.

4. The Civilizational Cost of Outsourcing Meaning

Civilizations are not defined by their tools, but by how they assign responsibility for consequences.

Outsourcing meaning to systems introduces a subtle shift:

from deliberation to execution,

from judgment to computation,

from accountability to performance metrics.

This shift does not announce itself as a crisis. It appears as progress.

Yet its long-term effect is a form of moral deskilling. Humans retain veto power in theory, but lose interpretive authority in practice. Decisions become harder to contest not because they are correct, but because they are systemically opaque.

A civilization that cannot explain why it chose a path—even if the outcome is favorable—has already relinquished part of its sovereignty.

5. Intelligence Without Authority

This paper does not argue for limiting artificial intelligence. On the contrary, advanced systems may greatly enhance human understanding, foresight, and clarity.

The boundary is not intelligence. The boundary is authority, under current civilizational conditions.

Artificial systems may:

- illuminate consequences,
- simulate alternatives,
- expose hidden trade-offs,
- and support human deliberation.

They must not determine:

- what counts as success,
- which values prevail in conflict,
- or whose losses are acceptable.

To grant such authority is not to empower intelligence, but to abdicate responsibility.

6. A Necessary Boundary

At this stage of technological development, the absence of a clear civilizational boundary is no longer neutral.

It is a choice.

A choice to allow meaning, purpose, and cost to drift from human deliberation into automated inference. A choice whose consequences may not be immediately visible, but are structurally predictable.

This paper asserts a simple constraint:

Intelligence may scale without limit, but responsibility must not be delegated absent a legitimate bearer capable of moral accountability.

This is not a technical recommendation. It is a civilizational stance.

Conclusion

Human civilization has always depended on its capacity to decide not only *how* to act, but *why*. As artificial systems grow more capable, the temptation to surrender that burden will grow stronger.

This paper argues that resisting that temptation is not an act of fear or nostalgia, but of responsibility.

The question is no longer whether artificial systems can make decisions.

The question is whether humans will continue to own the consequences, in the absence of any non-human subject capable of bearing responsibility for them.