The Invariant of Answerability

Luke F. Walton

Independent Researcher

ORCID: 0009-0005-9263-1954

Working paper · June 2026

Working paper. This note states the central claims and framing of a paper in preparation; the full development is forthcoming alongside its companion papers. It is circulated to place the claims on record.

Competing interests: The author is the founder of Surmado, Inc.; the account this paper assigns reaches the author's own layer.

Companion papers: *The Decision No One Authored* (the special case) and *The Captured Oracle* (the live demonstration).

Abstract

Interposing a machine between an agent and an outcome changes who must answer for it; it does not change whether an answer is owed. This paper defends that claim in its general form. The owing is anchored in whoever an action reaches, not in the route the action takes, and no mediation defeats it; only the party owed can release that account. The structure is old; the demand for an account from an externalized voice that cannot answer for itself is at least as old as writing. What is new is that the externalized voice through which we increasingly act no longer merely strands the demand for an answer, as a written text does, but answers back — fluently, in its own voice, while standing behind nothing. That single change is what converts a perennial structure into a pressing one and is the occasion for the paper.

The paper proceeds by distinguishing whether an account is owed from where it can be demanded: the first is invariant under routing; the second relocates, or fails, while the first holds throughout. It locates where an account comes to rest when a chain of authoring hands lies behind an outcome, and it does this without resolving whether a machine could itself ever become a party that answers — a question it holds open by design, on the same terms it holds open whether such a machine could be wronged. The claim is held constant across the metaethical space — true whether answerability is a brute normative fact or a constitutive feature of practices we cannot coherently abandon —

and is rested on neither. The contribution is the invariant and the allocation it forces; it is not a verdict on the machine's mind.

Opening

A companion to this paper earned a principle on a single channel and declined, rightly, to claim it everywhere. It studied what happens when an answer engine returns a verdict in its own voice with no author a reader can reach, and it named the constant beneath that case: answerability is conserved; interposing a machine changes who owes the answer, never whether one is owed. It stopped there, at the edge of one channel, and left the general claim unmade.

The general claim is old enough to predate every machine. At the founding of written argument, Socrates objects that a written text is like a painting: question it, and it returns the same words; it cannot defend itself, and it always needs its parent to come to its aid. The objection is not that the text lies. It is that the text strands the demand for an answer — issues a claim and leaves no one present to be asked. The demand does not thereby vanish; it stands, unmet, pointing at an author who is not there.

What has changed is the one thing that makes the ancient objection press rather than merely recur. The externalized voice no longer returns the same words and falls silent. It answers — composes for the occasion, defends when pressed, revises when corrected, in a fluent first person — while standing behind nothing it says. The text now talks back, and talking back is exactly the behavior we take, in each other, as the mark of someone who can be held. To take it as that mark here is the error the whole paper is built around. The demand for an account is met, in form, by a thing that cannot bear it; the account does not strand visibly, as the written page's did, but is answered away.

This paper makes the general claim the companion declined to make, and makes it on its own ground — not derived from the channel, which never needed the height, but defended directly: whenever an action reaches a party, an account is owed, and no routing of that action defeats the owing.

Claims

The following are the claims the paper defends, stated here on the record. The arguments by which each is established are developed in the full paper.

1. **The invariant, in general form.** Answerability is route-indefeasible: it is owed to whoever

an action reaches, it is undefeated by any mediation, and it is releasable only by the party to whom it is owed. (Where prior work establishes the invariance of responsibility under mediation, this claim wraps it; the contribution is its general form together with the structure below.)

2. **Owed versus dischargeable.** Whether an answer is owed is distinct from where it can be demanded. The first is invariant under routing; the second relocates, or fails. This distinction is what makes the conjunction *indefeasible yet relocating* a structural fact rather than a contradiction.

3. **Discharge can fail, while the owing holds across the failures.** The demand for an account can go unmet without the owing lapsing; the owing is what stays constant when discharge fails. This converts the observation that some cases have an obvious bearer into the claim the paper defends: that the invariant is precisely what remains while discharge fails, not a generalization from the cases where a bearer is ready to hand.

4. **The answers-back update is the urgency.** The shift from *stranding* (the written text that issues a claim and leaves no one present to answer for it) to *fluent, unanswerable response* (the externalized voice that composes, defends, and revises while standing behind nothing) is what converts an ancient structure into a pressing one. This is the paper's account of why the problem presses now, and its case that the claim is not trivial.

5. **The metaethical posture.** The invariant is held constant across the whole metaethical space and rested on no part of it: the same discipline the program applies to whether a machine can be wronged and whether it can come to answer for itself.

The metaethical posture

This paper does not turn on whether answerability is a brute normative fact about the world or a constitutive feature of practices a functioning society cannot coherently abandon. The distinction is real and is argued elsewhere; it does not bear here. The claim is built to hold across the whole range it opens — at either pole and at every mixture — and to rest on no part of it. This is not agnosticism deployed to avoid a fight one cannot win. It is the narrower and more demanding posture of holding a thing constant without holding it foundational: showing that the truth of the invariant does not co-vary with where a reader lands on a question the author is not competent to settle, and that no one yet is. The same posture governs the paper's two other refusals — whether the machine can be wronged, and whether it can come to answer for itself — and naming the posture, rather than smuggling it, is part of what the paper contributes.