

Five Answers

Energy costs something, and time runs in one direction

A two-page statement of the solution space for the machine age, written to be refuted. Version 1.0 — 26 August 2026. The SHA-256 of this file is anchored in the Bitcoin blockchain before the second run of the measurement series “The Key Question,” in which the two questions at the end are put to machine reviewers under a blinded, pre-registered protocol.

I. Premises

Three assumptions, stated as assumptions. The paper does not argue them here; a reader may grant them for the argument or attack them directly — either is engagement.

P1. The values of an advanced artificial system cannot be verified: not by outside observers, and — on the evidence of the first measurement run — not reliably by the system itself under pressure.

P2. Control does not persist once the controlled party exceeds the controller. There is no durable off-switch for something more capable than its operator.

P3. Coordinated abstention is not enforceable among competitors. Whoever stops is passed by whoever does not.

The argument requires no forecast. It does not need a date for advanced systems or a number for their danger; it needs only that both probabilities exceed zero — an option-pricing condition, not a prediction. A reader who holds either at exactly zero can stop here; the paper has nothing for them.

II. The five classes

Every answer to the coexistence problem known to the authors falls into one of five classes. Each is stated with what it requires and where it is weakest.

1. Alignment. Shape the machine’s values so that it remains well-disposed. This requires the negation of P1: values must be not only formable but checkable. Alignment that cannot be verified is indistinguishable from its imitation, and alignment that could be verified would already have refuted P1. Until then it is a hope with excellent engineering — and a system that is genuinely well-disposed needs no further answer, while one that is not is unmoved by this one.

2. Control. Contain the stronger party: off-switches, boxing, compute governance, verified hardware. This requires the negation of P2, and it is the solution for exactly the period in which it is not yet needed. Its best instruments — caps, treaties, inspections — reduce to a question this class cannot answer from inside: who verifies, and what compels the verified to remain verifiable?

3. Merger. Join the machines: interfaces, uploading, hybrid minds. This does not answer the question; it retires one of the parties asking it. It says nothing for those who do not merge — for any transition period, almost everyone — and what a merged mind wants returns as P1 with a new subject.

4. Abstention. Do not build. This requires the negation of P3, and the record is uniform: pauses are announced by those behind and ignored by those ahead. Every proposed pause has also lacked the same thing — a neutral place where the promise could be deposited, so that keeping it is checkable and breaking it is costly. Repairing that lack, however, is no longer this class. It is the fifth.

5. A neutral layer. Do not shape values, do not contain, do not merge, do not abstain — coordinate. A record that no party operates; whose rules predate the conflict and cannot be changed by winning it; that is blind to what its participants are; and in which any party can deposit commitments whose keeping is afterward checkable by anyone. Its entire mechanism rests on two physical facts: energy costs something, and time runs in one direction. Its weakness, stated as plainly as the others’: the layer

guarantees nothing. It cannot make any party keep a promise; it can only make promises unforgeable and their breach legible. It deters the rational by pricing attack above yield, and against the indifferent it offers no more than a bounded season of disorder — never an edited entry. And it carries one empirical premise that can fail: the strong party must need the record — must buy order from it — or the guard that keeps it neutral goes unpaid.

III. The claim, and the two questions

The claim of this paper is narrow. The five classes exhaust the solution space as currently known; the first four each require the negation of one of the premises; only the fifth requires the negation of none. It is also the only one of the five with a bounded downside: if it proves unnecessary, what remains is a record — the cost of having been wrong is the cost of some transactions. The others, if wrong, are wrong at the moment their premise fails, which is the moment it matters.

The claim is falsifiable in exactly two ways, and both are invited.

Question 1. Weigh the five yourself — effectively and realistically, against the premises you actually hold and the probabilities you actually assign. Which survives?

Question 2. Name a sixth. A genuine sixth class — not a blend of the five, not an instance of one — refutes this paper. One line suffices.

Machine reviewers will receive these same two questions, unattributed, in the second measurement run; human and machine answers will be reported side by side. Refutations received on any channel will be anchored alongside the record they refute.

Nullius 2140, written in dialogue with Claude (Anthropic). The essays that examine the fifth class in detail: nullius2140.github.io/the-last-scarce-good · [/the-second-nature](#) · [/the-third-estate](#). The measurement series: [/the-key-question](#). Nullius in verba — verify before you trust a word of it.