

Unauthorized Futures: Possibility, Demonstration, and the Limits of Retroactive Vindication

Flyxion
Independent Researcher

October 2026

Abstract

Charlie Warzel's October 2026 commentary on the Elizabeth Holmes documentary *You Can See Everything*, read alongside the Tesla Roadster unveiling and the present artificial intelligence boom, describes a culture in which claims about technologies that do not yet exist are treated as grounds for present commitments. This essay isolates the epistemic structure of that culture. It argues that three propositions are routinely collapsed: that a technology is possible, that it is reachable from present conditions, and that it is sustainable once reached. Each of these is a separate claim requiring separate evidence, and none of them settles the further question of whether a particular present action is authorized. The essay develops a staged account of how a claim acquires standing to govern a decision, shows formally that later success cannot retroactively supply the ex ante epistemic warrant that an earlier authorization lacked, and shows that unbounded promises of eventual arrival cannot be refuted by any finite run of failures. It then examines the incentive structure that rewards overstatement when the penalty for small exaggerations is negligible, the heterogeneity of the artificial intelligence field as a source of equivocation, and the neglected problem of maintenance. The argument is not directed against speculation, which is indispensable to research and finance, but against the conversion of speculation into authority without the evidence appropriate to the commitment requested.

1 Introduction

A commentary broadcast by Charlie Warzel on the *Galaxy Brain* program of *The Atlantic* places three events within a single frame. The first is a public demonstration by Tesla of a prototype Roadster that climbed a partial loop, hung inverted, and returned to the ground using compressed air thrusters. The second is the release of a documentary about Elizabeth Holmes, the convicted founder of Theranos, and her partner Billy Evans, who is shown trying to assemble what they call a second version of the company. The third is the condition of

the contemporary artificial intelligence industry, in which claims about future capability are used to justify enormous present investment. Warzel’s thesis is that these are connected by an ethos in which the figure of the builder, a high-agency visionary who is never wrong but only early, has become more closely identified with persuasion than with product.

The commentary is journalism, and its generalizations about an entire industry are offered as observations rather than as the output of a systematic study. That does not diminish its usefulness as a point of departure. The scenes it selects display a recurring epistemic pattern with unusual clarity, and the pattern can be stated precisely enough to be analyzed. The pattern is that a proposition about a future state of affairs is allowed to function as a premise in decisions made now, by parties who cannot inspect it and who bear costs if it fails. The closing remark of the commentary, in which an investor and podcast host predicts that within a decade a blood-testing product will show that Holmes was merely twenty years too early, compresses the pattern into a single sentence. If the technology eventually exists, the earlier claim is held to have been vindicated.

The aim of this essay is to take that sentence seriously as an inference and to determine what it can and cannot establish. The argument proceeds in several steps. The next section separates three notions of future viability and a fourth question about authorization. The third section gives a staged account of standing and proves a non-retroactivity result. The fourth examines why promises of eventual arrival resist refutation. The fifth turns to the incentives of venture finance and the allocation of downside risk. The sixth considers the artificial intelligence field, where legitimate demonstrated capability and unsupported extrapolation coexist under one name. The seventh addresses maintenance, which Warzel raises through Elizabeth Renieris’s notion of imaginative obsolescence. The eighth proposes a diagnostic for distinguishing hypotheses from commitments, and the final section states limits.

2 Possibility, reachability, sustainability, and authorization

Public argument about technology typically moves between claims of very different logical strength without marking the transitions. It is useful to separate them. Let F denote a described future state, for example a handheld device that performs a wide range of blood tests from a finger-prick sample, a road vehicle capable of controlled flight, or a population of software agents performing most economically valuable cognitive work. Four questions can be asked about F .

The first is whether F is possible, written $\text{Poss}(F)$. This is a question about whether physical law and the structure of the relevant domain permit the state. It is answered by theory, by existence proofs in related settings, and by the absence of a principled obstruction. A positive answer is weak. Many states are possible in this sense and will never be realized.

The second is whether F is reachable from the present, written $\text{Reach}(F \mid X_0, \mathcal{C})$, where X_0 is the current state of knowledge, materials, institutions, and capital, and $\mathcal{C}$ is the set of

constraints that any construction path must respect. Reachability requires an admissible trajectory γ from X_0 to F , that is, a sequence of steps each of which is individually achievable within the constraints of energy, cost, time, supply, regulation, and available expertise. Describing a destination is not the same as exhibiting a path. A founder who describes a future in which a tiny blood sample yields hundreds of results has specified an endpoint. The existence of a biochemical, manufacturing, and clinical path to that endpoint is a separate matter, and it was precisely that matter that Theranos could not support.

The third is whether F is sustainable, written $\text{Sust}(F \mid \mathcal{C}')$. A state that can be reached once, under demonstration conditions, may not be maintainable as a continuing operation. The constraints $\mathcal{C}'$ governing continued operation differ from those governing construction, since they include maintenance, replacement of components, error correction, supply continuity, and the persistence of the institutional knowledge required to keep the system functioning. A prototype that performs a maneuver once under controlled conditions has demonstrated a fragment of reachability and almost nothing about sustainability.

The fourth question is of a different kind. It asks whether a particular present action a , such as an investment, a clinical deployment, a public statement of capability, or the redirection of an economy toward the anticipated technology, is authorized, written $\text{Auth}(a \mid E_t)$, where E_t is the evidence available when the action is taken. Authorization is not a property of F alone. It depends on what is being asked of whom, on how reversible the action is, and on who bears the cost if the claim proves false.

These four predicates are distinct, though they are not wholly unrelated. Under the natural reading, a state that is reachable by an admissible trajectory is thereby possible, so that $\text{Reach}(F \mid X_0, \mathcal{C}) \Rightarrow \text{Poss}(F)$, provided that both predicates refer to the same domain and that the constraints are mutually consistent. Sustainability, in the sense of continued operation of a state that has actually been attained, presupposes reachability in turn, although a weaker reading, on which it concerns only the existence of a stable operating regime, does not. The converses all fail. A possible state may be unreachable, a reachable state may be unsustainable, and a state that is possible, reachable, and sustainable may still be a state that no particular action is entitled to presuppose on present evidence, since none of the three predicates by itself establishes Auth . The central error that Warzel's examples display is the use of the weakest of these, usually an assertion of possibility supported by charisma and an appeal to inevitability, to underwrite the strongest, an authorization to commit resources or to affect people who did not agree to participate in an experiment.

The Roadster demonstration illustrates the gap between the first three questions without any suggestion of bad faith. According to Warzel's account, the demonstrated vehicle used ten thrusters powered by roughly six hundred liters of compressed air, the production vehicle is not expected before 2028, and the flying version is intended for invite-only sale at a likely very high price. The maneuver shown was real. What it established is that a specially configured prototype can perform a maneuver under controlled conditions. It did not establish road legality, operational safety, or manufacturability at scale, and Warzel's

observation that he was asked to treat the pledge to deliver as an article of faith identifies exactly the point at which a demonstration is being asked to carry more than it can bear.

3 Standing and the staged conversion of claims into commitments

A claim does not govern a decision merely by existing. It must pass through stages, and the stages can be separated conceptually even where institutions compress them. A claim is first generated by someone. It is then inscribed in some durable and inspectable form, so that others can examine what exactly was asserted. It is then evaluated against evidence appropriate to its content. A decision procedure then determines whether the claim is authorized to govern a particular action, and only after that does an irreversible commitment occur. Writing these as $G \rightarrow I \rightarrow E \rightarrow A \rightarrow C$ for generation, inscription, evaluation, authorization, and commitment, the point of the staging is that each transition requires conditions that the preceding stage does not supply. A generated claim has not been inscribed in an inspectable form. An inscribed claim has not been evaluated. An evaluated claim has not thereby been authorized for every purpose. The sequence is an analytical ordering and not a claim about institutional chronology. In practice the stages are often entangled, since inscription can itself be a form of evaluation when a disclosure requirement forces a claim into checkable terms, and evaluation is frequently shaped by what the authorizing body is willing to accept. The ordering is defended on a narrower ground, namely that each stage answers a different question (what exactly was claimed, whether the evidence supports it, whether that support suffices for this action, and whether the action is now irreversible), so that the collapse of any two of them can be identified by asking which of the four questions was never put.

The standard of evaluation that is adequate depends on the commitment requested. Clifford's *The Ethics of Belief* insisted on the moral importance of evidence and on the social consequences of credulity, and it took a position more categorical than the one developed here (Clifford, 1877). The graded threshold introduced below is offered as a decision-theoretic extension of his concerns and not as a principle he stated: the strength of the grounds required for acting on a belief is taken to increase with the consequences of being wrong. A research grant that funds an uncertain hypothesis commits resources to an acknowledged conjecture and needs a correspondingly modest evidential threshold, since everyone involved understands that the object of the investment is the investigation. A medical test sold to patients commits those patients to decisions about their health on the strength of the reported result, and the threshold is correspondingly demanding. The same sentence, that this technology will work, can be adequate for the first purpose and wholly inadequate for the second.

Let $\theta(a)$ denote the evidential threshold that an action a requires, increasing in the irreversibility of a and in the extent to which its costs fall on parties who did not accept

the risk. Let $W(C \mid E)$ denote the degree of support that evidence E lends to claim C , on whatever ordinal scale the relevant domain supports. No numerical calibration is assumed here, and the schema carries a simplifying idealization that should be stated openly: it treats warrant as totally ordered and treats $\theta(a)$ as comparable with it, so that the two can be set side by side. Real evidence is often multidimensional, with direct measurement, mechanism, and track record supporting a claim in incommensurable ways, and real thresholds are set by negotiated institutional rules and not by a single function. The condition is offered as a schema that makes the structure of the argument visible, and the results below use only its monotonic form, namely that more irreversible or more externalized actions require at least as much support as less irreversible or less externalized ones. The authorization condition is then

$$\text{Auth}(a \mid E_t) \iff W(C \mid E_t) \geq \theta(a),$$

where C is the claim on which the action depends.

The pathology that Warzel describes can now be located. It is the movement directly from G to C , from the generation of a compelling possibility to a consequential commitment, with the intermediate stages bypassed or replaced by reputation, demonstration, enthusiasm, or assertions of inevitability. Theranos is the extreme case, in which the claim was inscribed in forms that concealed rather than disclosed what the system could do, and evaluation was obstructed rather than performed. The more common and subtler case is one in which every stage is formally present but the evaluation is conducted by parties who share the promoter's incentives, so that the stage exists in name and not in function.

3.1 Non-retroactivity of justification

The closing prediction in the commentary, that a future blood-testing product will show Holmes to have been early rather than wrong, relies on an inference that can be made explicit. Write T for the technology and C_t for the claim, made at time t , that T works as represented. The defense reasons that if T is eventually realized, then the claim and the authorizations that depended on it are vindicated:

$$T \text{ is realized at some } t+k \implies \text{Auth}(a \mid E_t) \text{ was correct.}$$

Three temporal judgments should be kept apart. The first is the ex ante epistemic justification of the action, $J_t(a)$, which holds exactly when $\text{Auth}(a \mid E_t)$ does, so that it is fixed by the evidence available at t . The second is a later assessment of the truth of the claim, $T_{t+k}(C_t)$, made on E_{t+k} . The third is a later legal or institutional assessment of the action, $L_{t+k}(a)$, which may take account of outcomes, of newly available evidence, and of standards that were not themselves in force at t . Later evidence can change the second and may change the third, but it cannot change the first, because it cannot be made available retroactively to the person who decided. The proposition therefore concerns ex ante epistemic justification specifically, and makes no claim about every sense in which an

action may later be assessed.

Proposition 1 (Non-retroactivity). *Let a be an action taken at time t on the basis of claim C_t , and suppose $W(C_t \mid E_t) < \theta(a)$. Let $E_{t+k} \supseteq E_t$ be the evidence available at a later time, and suppose $W(C_t \mid E_{t+k}) \geq \theta(a)$. Then $J_t(a)$ fails, and it continues to fail however strongly E_{t+k} supports C_t .*

Proof. Ex ante justification at time t is defined as a function of E_t , the evidence available to the decision at that time. The hypothesis states that $W(C_t \mid E_t) < \theta(a)$, so $\text{Auth}(a \mid E_t)$ fails, and with it $J_t(a)$. The quantity $W(C_t \mid E_{t+k})$ is evaluated on evidence that was not available to the decision. Its value bears on $T_{t+k}(C_t)$, the later question of whether the claim turned out to be true, but $J_t(a)$ does not take it as an argument. A justification of this kind is a relation between a decision and the information on which it could have been made, and enlarging the information afterward does not change what the decision could have been made on. $\square$

The proposition separates outcome from warrant. A decision can be lucky, in the sense that its premise turns out to be true, without having been justified. Conversely, a decision can be well justified and turn out badly. This is an ordinary distinction in the theory of rational choice, and its application here carries a specific consequence. It means that the later existence of a technology cannot repair the earlier misrepresentation of it. If a company in 2015 represented that its device performed a test that the device could not perform, and another company in 2035 develops a device that performs it, the 2015 representation remains a misrepresentation, and the reliance placed upon it by investors and patients remains reliance on an unsupported statement.

The distinction matters particularly for the legal and moral status of the claim. Warzel’s description of Holmes’s conviction as concerning fraud, together with the broader historical record of unreliable testing, illustrates that legal findings, engineering failures, and ethical judgments are related but separate categories. The conviction concerned representations to investors. The ethical significance of patient exposure to unreliable results is a further matter. A later technological success can bear on whether the underlying goal was achievable and may shape some later institutional assessments, but it leaves the ex ante epistemic justification of the original decision exactly where it was.

4 Eventual arrival and the insulation from refutation

The phrase that a builder is never wrong, only early, has the logical form of an existential claim over an unbounded horizon: there exists a time at which the technology will exist. Such a claim is not refuted by any finite sequence of failures. If the product does not exist in year five, the claim is untouched, since the arrival may occur in year six. If it does not exist in year fifty, the claim remains untouched for the same reason. The claim is safe from

falsification not because it is well supported but because its logical form prevents finite evidence from bearing on it.

The closest fit is Lakatos’s account of research programs, in which a program is progressive when its modifications predict novel facts and degenerating when its modifications serve only to protect the core from adverse results, and the pattern at issue is a degenerating problem-shift in which each missed date is absorbed by a revised date (Lakatos, 1970). Popper’s criterion of falsifiability supplies the background against which the pattern is visible, since it identified insulation from refutation as a mark of claims that cannot be tested (Popper, 1959), but his criterion is blunter than the case requires. The contrast is directly relevant. A research program in a healthy science may certainly persist through long periods of apparent failure, and Lakatos’s point was that persistence can be rational. What distinguishes a legitimate persistence from a degenerating one is whether the program commits to consequences that could, in principle, count against it, and whether its adjustments generate new content or merely absorb refutation.

Let ρ denote an interpretive map that assigns to each observed outcome o a verdict on whether it counts for, against, or neutrally with respect to the claim C . The claim is protected from refutation when $\rho(o) \in \{\text{for, neutral}\}$ for every outcome o that could be observed. Failure is read as prematurity, delay as the cost of ambition, and discrepancy as the market not yet being ready. Under such an interpretive map the claim is protected from every outcome, and nothing that could be observed would discriminate between the promoter being right and the promoter being wrong. This is a stronger condition than the mere absence of a finite decisive failure. An unbounded prediction can lack such a failure condition and still be responsive to evidence, since intermediate achievements bearing on the mechanisms expected to produce the outcome can legitimately raise the credibility of eventual success without proving it, and repeated missed indicators can lower it. The logical point about falsification must therefore be kept distinct from the point about Bayesian updating: a claim can be non-falsifiable in the strict finite-history sense while remaining more or less credible given evidence about mechanism. The pathology lies in the conjunction of an unbounded horizon with an interpretive map under which every outcome is read as favorable.

Proposition 2 (Insulation by unbounded horizon). *Let C be the claim that a state F will be realized at some time in an unbounded future. Then for any finite observation history h in which F has not yet been realized, h is logically consistent with C .*

Proof. Realization of F at a time later than the end of h is consistent with h by construction. Since the horizon is unbounded, such a time always exists. Therefore no finite h entails the negation of C . $\square$

The remedy is not to forbid long-horizon claims, which are indispensable in research and infrastructure, but to index them. A claim of the form that F will be realized by time τ , under stated conditions, and with stated interim indicators, can fail, and its failure is

informative. A commitment that depends on such an indexed claim can be proportioned to the interim evidence, released in stages, and withdrawn when indicators are missed. Warzel’s account of the cultural environment suggests that the unindexed form is both more rhetorically powerful and more common in the contexts he describes, since it is the form that supports the claim that the permanent loss of economic standing awaits those who do not commit early.

5 Venture finance and the allocation of downside

Warzel attributes a considerable part of the culture to the structure of venture capital, which deliberately purchases exposure to uncertainty in the expectation that a small number of exceptionally large outcomes will compensate for many failures. That structure is not irrational. For an investor considering ventures $i = 1, \dots, n$ with costs c_i , success probabilities p_i , and payoffs R_i , the expected return of the portfolio is

$$\mathbb{E}[\Pi] = \sum_{i=1}^n (p_i R_i - c_i),$$

and it is possible for an individual venture to have a very low p_i and still be a rational investment when R_i is sufficiently large and the investor’s exposure is bounded by c_i . The knowing acceptance of such risk by a party who has the information and the capacity to absorb the loss is not a failing.

The difficulty arises with respect to who is exposed. The investor in an uncertain medical technology has accepted the risk. A patient who receives a misleading test result has not. The asymmetry is not about the existence of speculation but about the identity of the party authorized to speculate relative to the party who bears the consequence of failure. When upside is captured privately and downside is transferred to workers, customers, public institutions, or the broader economy, the speculation is being conducted at the expense of parties who had no occasion to price it.

A second structural feature concerns who receives the reward for overstatement and who pays for it. Let a promoter have a true capability k and advertise a capability $s \geq k$. Suppose the reward from advertising is an increasing function $R(s)$, reflecting that investors, customers, and talent are drawn to larger claims. Suppose the expected loss to the promoter or the promoter’s organization on discovery is a function $\Pi(s - k)$ of the size of the gap, and let $H(s - k)$ be the expected external harm to third parties such as patients, customers, and workers, with $\Pi(0) = H(0) = 0$. Let $\lambda \in [0, 1]$ be the share of that external harm that the promoter internalizes, through liability, reputation, or conscience. The promoter’s expected net reward is then

$$U(s) = R(s) - \Pi(s - k) - \lambda H(s - k).$$

Ordinary commercial practice tolerates small exaggerations, often described as puffery, so

that the marginal organizational penalty near zero is small, and the external harm of a small exaggeration is likewise small.

Proposition 3 (Overstatement under negligible marginal cost). *If $R'(k) > \Pi'(0) + \lambda H'(0)$, then advertising exactly the true capability $s = k$ does not maximize the promoter's expected net reward. In particular, this holds whenever $R'(k) > 0$ and $\Pi'(0) = \lambda H'(0) = 0$.*

Proof. Differentiating at $s = k$ gives $U'(k) = R'(k) - \Pi'(0) - \lambda H'(0) > 0$. Since U is differentiable and strictly increasing at $s = k$, there exists $s > k$ with $U(s) > U(k)$. $\square$

The distinction between the parties matters. The social net reward from the same behavior is $W(s) = R(s) - \Pi(s - k) - H(s - k)$, and the two quantities coincide only when $\lambda = 1$. When λ is small, the condition of the proposition can hold even where $H'(0)$ is substantial, so that the incentive to exaggerate persists although the external cost of exaggeration is large. This is the same asymmetry noted above between the party authorized to speculate and the party who bears the failure, now stated as an incentive condition, and it connects the argument to the question of non-extraction without requiring the adoption of any larger framework.

The result is elementary, and it depends on assumptions that can be examined. Where the marginal cost to the promoter of modest overstatement is negligible and the marginal reward for greater claims is positive, accuracy is not a stable equilibrium. This is not a claim that promoters act on such calculations consciously, only that an environment with this structure selects for the behavior over time, as a result of which the practice becomes ordinary and is described, in a remark Warzel quotes from Nikita Bier, as excused by adults in the room as faking it until making it. The proposition also indicates where an institutional response would act: on the penalty function, by making the cost of overstatement rise from the first increment, on the internalization parameter λ , by making promoters answer for external harm, and on the detection process, by making the claim inscribed in a form that can be checked.

This also states a research question about organizations. If an organization can obtain greater reward by exaggerating its future capabilities than by accurately reporting its present limitations, then epistemic distortion is not an occasional moral lapse but an economically stable strategy. The organization then has reasons to maintain uncertainty, to discourage verification, and to postpone accountability, and these reasons operate whatever the sincerity of its individual members.

6 A heterogeneous field under one name

Warzel is careful to say that chatbots and coding agents are real and powerful in ways that earlier speculative technologies, such as the metaverse or non-fungible tokens, were not. That qualification is essential and it creates the analytical difficulty. The field called artificial intelligence contains systems that have demonstrated capabilities, empirical measurements

of improvement, speculative research programs, and unsupported predictions, and public discussion frequently moves among them as though they were a single category with a single evidential status.

It is useful to distinguish four levels of claim. An observed capability is a performance on a defined task under described conditions, such as a coding agent completing a benchmark task. An extrapolation projects a trend in such measurements into the near future under an assumption that the trend continues. A conjecture concerns an architecture, a method, or a capability that has not been demonstrated at all. An economic forecast makes claims about adoption, substitution, regulation, and institutional change that depend on the previous three and on additional assumptions of their own. The evidential requirements differ at each level. An observation is supported by measurement. An extrapolation is supported by the stability of the trend and by an account of what could interrupt it. A conjecture is supported, if at all, by theory and by analogy. A forecast is supported by models whose assumptions can be inspected and varied.

The inference that Warzel criticizes most effectively is the one from the first level to the fourth. That a system completes a benchmark task is an observation. That such systems will transform the labor market is an economic hypothesis that does not follow from the observation without further assumptions about reliability, cost, adoption, liability, and organizational change. His example of a forecast that United States output could rise from roughly thirty trillion to forty-four trillion dollars within four years, described as resting on the most extreme parameter setting of a published economic model, demonstrates how a scenario can be converted into a headline by dropping the conditions that define it. The arithmetic is simple. A rise from thirty to forty-four over four years corresponds to a compound annual growth rate of

$$g = \left(\frac{44}{30}\right)^{1/4} - 1 \approx 0.1005,$$

or approximately ten percent per year. Whether that figure is extraordinary depends on whether the quantities are nominal or real, on whether the statement is a forecast or a conditional scenario, and on the baseline against which it is measured. The statement, as reported, does not settle those matters, and a reader cannot assess it without them. The point is not that the scenario is false but that its evidential status is that of a conditional model output, and that presenting it without the conditions turns a modeled possibility into an apparent expectation.

The levels interact in a way that makes the slide from one to the next easy to miss. Each level inherits the uncertainty of the one below and adds assumptions of its own, so that confidence in the top level can be no greater than confidence in the whole chain, and a forecast that stacks a continued trend, an undemonstrated method, and a favorable adoption path is weaker than any of its parts. Public claims can be placed on this ladder. A coding agent completing a benchmark task sits at the first level. The expectation that such performance

will continue to improve at its recent pace sits at the second. The statement that autonomous agents will perform most economically valuable cognitive work sits at the third where it depends on capabilities not yet demonstrated, and at the fourth where it is converted into a statement about output, employment, or valuation. The Warzel commentary is at its most persuasive on exactly this point, since the output forecast and the valuation comparison are fourth-level claims presented with the confidence that is warranted, at most, for the first. A sound public statement would say at which level it is speaking and what the levels beneath it contribute, and the authorization of large commitments should be tied to the level at which the supporting evidence actually sits.

A related comparison concerns the stated ambition that a company’s valuation might come to exceed the current world economy by orders of magnitude. A valuation is a price for a claim on future cash flows, and annual economic output is a flow of production, so the comparison places quantities of different kinds in one ratio. It can be rhetorically striking without carrying a financial forecast. In the vocabulary of the earlier sections, it is a statement of $\text{Poss}(F)$ presented in a form that invites the reader to infer $\text{Reach}(F)$.

Treating all of this as a single phenomenon with Theranos also reproduces the confusion in question. Theranos was a case of materially misleading representation of a capability that did not exist as described. A contemporary system with a demonstrated capability that is described accurately is a different case, and a description that conflates the two is itself an equivocation, however well intended. The sound form of the critique distinguishes the levels and applies the corresponding requirement to each, rather than treating the whole field as a bubble or the whole field as vindicated.

7 Maintenance and the political economy of continuation

Warzel quotes Elizabeth Renieris’s phrase imaginative obsolescence for a condition in which one vision of the future so rapidly replaces the next that the systems already in place decay. The phrase identifies a feature of the sustainability question raised earlier. A civilization consists not only of innovations but of systems that must go on operating after they are built. Bridges require inspection, software requires patching, medical institutions require reproducibility of results, data infrastructure requires stewardship, and institutional knowledge requires transmission to those who come later. A culture organized around novelty undervalues these activities, because maintenance does not generate the spectacular returns that justify breakthrough investment, and because its successes are invisible: the bridge that does not fall is not an event.

Local improvements do not necessarily compose into global improvements. Let V_i be a measure of the viability of the i th component of a system and V_{glob} a measure of the viability of the whole. It is entirely possible that

$$\Delta V_i > 0 \quad \text{while} \quad \Delta V_{\text{glob}} < 0,$$

since a component can increase its own performance by drawing on shared resources, by making other components dependent on it, or by shifting costs to parts of the system that are not measured in the local objective. This is not a paradox. It states that the local objective is not a faithful representation of the global constraints, and that a technology evaluated only by the performance of its isolated component has not yet been evaluated for what matters to continued operation. A computational advance that increases energy demand, reduces the security of the labor on which other services depend, or makes essential services dependent on a single provider may be an advance in one register and a loss in another.

The relation to the earlier argument is direct, and it can be stated in terms of the authorization condition. Evidence about sustainability is of a distinctive kind, because it can only accumulate through continued operation over time and cannot be supplied by a demonstration, which by design shows a state at a moment. A deployment whose threshold $\theta(a)$ includes the requirement that the system keep working therefore cannot be authorized on demonstration evidence alone, however impressive the demonstration. A technology that is claimed to be possible has not been shown to be sustainable, and the question of who will maintain it, at whose cost, and with what continuity of knowledge is a question that promotional claims about arrival rarely address. The claim here is limited to promotional discourse, where the omission is close to structural, and not to every sector, since mature engineering fields treat maintenance as a first-order design constraint. A system that rewards creation while externalizing the cost of maintenance may increase its apparent productive capacity while diminishing its capacity to persist. This is the political economy of continuation, and it may be the most consequential of the themes in the commentary, since it concerns not whether particular claims are true but whether the surrounding culture can sustain the things it has already built.

8 Correction and the diagnostic of evidence-sensitivity

The recurring observation in the commentary is that failure does not appear to alter commitments. A claim that has been shown to be wrong is redescribed as premature, and a failure is redescribed as an experiment. The relevant distinction is between a repair that changes the organization’s model in response to evidence and a repair that changes the interpretation of the evidence in order to preserve the model.

Let M_t denote an organization’s model of its own capabilities at time t , and let e_t be new evidence. A corrective process produces $M_{t+1} = \mathcal{R}(M_t, e_t)$, where $\mathcal{R}$ is a repair operator. Let $\mathcal{E}$ be an evaluation protocol, which specifies which observations count and how discrepancy between what a model predicts and what the evidence shows is measured, and let $D_{\mathcal{E}}(M)$ denote the discrepancy of model M under it. A corrective operator is one for which, under stated conditions, $D_{\mathcal{E}}(M_{t+1}) \leq D_{\mathcal{E}}(M_t)$, where $\mathcal{E}$ is held fixed across the comparison and is independent of the organization being assessed, or where any change to $\mathcal{E}$ is independently

justified. The qualification is essential, because a system can reduce a measured discrepancy by changing the measurement function, excluding inconvenient observations, or redefining the target, and an inequality stated without a fixed protocol cannot distinguish correction from narrative preservation. Even with the protocol fixed, the inequality is an illustrative condition for correction and is not offered as a theorem. The pathological alternative is an operator whose objective is the preservation of a narrative, in which the model is held fixed and the mapping from evidence to verdict is adjusted until the discrepancy disappears. This can be called anti-repair, because it has the form of adaptation while preserving the object that the evidence called into question.

The diagnostic that follows is practical and does not depend on any judgment about the sincerity of the persons involved. For any claim on which a commitment depends, one asks what evidence would cause the organization to withdraw the claim, alter its commitments, compensate affected parties, or abandon the project. If an answer can be given, and the evidence is of a kind that could in fact be obtained, then the claim is functioning as a hypothesis. If no conceivable result would trigger correction, then the claim is no longer functioning as an empirical hypothesis. It has become an organizing commitment protected against evidence.

The commentary adds a psychological consideration in noting that Holmes's identity appears inseparable from being a builder, so that for her a revision of the claim might be a revision of the self. The documentary's structure, which Warzel describes as burying the audience in unreliable narration, makes it difficult to infer any subject's actual beliefs from filmed behavior, and the interpretation should be treated as his reading and not as an established psychological fact. Granting it as a hypothesis, however, it points to an institutional lesson that does not depend on its truth. If a founder's identity is tied to the correctness of a founding claim, then the founder has two objectives that need not be compatible, accuracy about the world and the stability of the self, and an organization that equates its mission with its founder's identity will carry that tension into its procedures. A trustworthy institution therefore cannot rely on the capacity of its founder for personal self-correction. It must be able to correct its claims even when its founder cannot tolerate the correction, which means that the procedures for inscription, evaluation, and withdrawal must be located outside the person whose identity is at stake.

9 Limits of the argument

Several limits should be stated plainly. First, the empirical premises about the cultural condition of the industry, including the decline of institutional trust, the influence of industry-friendly media, and an increasingly permissive attitude toward speculative promises, are plausible generalizations that the commentary advances without systematic longitudinal evidence. Showing that some founders exaggerate does not show that exaggeration has become more common. Establishing that would require a defined population of claims,

consistent classification criteria, and comparable evidence across time. The counterfactual about how the industry would respond to a collapse like that of Theranos if it occurred today is a thought experiment and should remain one.

Second, several of the numerical and quotational assertions in the source are reported by the commentator and have been treated here as his account, without independent confirmation of the underlying figures. This applies to the attributed forecast of output growth, to the stated size of metaverse investment, to the statement about the scale of the documentary's opening, and to the quotations from named individuals. The argument of the essay does not depend on their precise values, but a reader who wishes to rely on any of them should consult the primary sources.

Third, the formal propositions above are deliberately modest. The non-retroactivity result is close to a definitional consequence of indexing justification to the information available at the time of decision. The insulation result is an observation about logical form. The overstatement result depends on assumptions about the reward and penalty functions that are plausible but empirical. Their value lies in making the structure of the argument explicit so that its assumptions can be challenged, and not in supplying a quantitative model of any particular firm or industry.

Fourth, the distinction between hypothesis and commitment does not condemn risk. Speculative research, early-stage finance, and ambitious engineering are legitimate and often socially valuable, and many successful inventions have come from organizations that made claims well in advance of the evidence. What the analysis condemns is narrower: the use of a representation of the future to secure authority over present resources and persons without satisfying the evidential requirement appropriate to the commitment requested, and the subsequent insulation of that representation from the evidence that would otherwise correct it.

10 Conclusion

The commentary's closing image is of a builder who is never wrong but only early, to be vindicated when the technology finally arrives. The analysis of this essay indicates why the image is attractive and why it fails as an epistemic principle. It is attractive because it is true of some cases: some ideas are premature, and some founders who were derided did turn out to be right. It fails because it cannot distinguish those cases from the others. A claim of eventual arrival, in its unindexed form, is compatible with every finite history, and so cannot distinguish the premature from the mistaken. Later success, in the rare cases in which it comes, does not supply the warrant that an earlier authorization lacked, since justification belongs to the evidence available when the decision was made.

Possibility, reachability, sustainability, and authorization are separate questions, and the discipline that the present moment appears to require is the habit of asking which of them a given claim has actually answered and which it has merely implied. The corresponding

institutional requirement is that claims be inscribed in forms that others can inspect, evaluated by parties without a stake in the answer, indexed to horizons and indicators that can fail, and tied to commitments that can be withdrawn. Under such conditions speculation can continue to do what it does well, which is to direct attention and resources toward uncertain goods, without being converted into an entitlement. The future may vindicate a claim, but it cannot retroactively authorize a deception.

References

- Clifford, W. K. (1877). The ethics of belief. *Contemporary Review*, 29, 289–309.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Knowledge* (pp. 91–196). Cambridge University Press.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson.
- Warzel, C. (2026). The faith-based logic of Silicon Valley. *Galaxy Brain, The Atlantic*, October 2026. Transcript of a commentary on the documentary *You Can See Everything* and the Tesla Roadster demonstration. Quotations attributed in the text to Nikita Bier, Elizabeth Renieris, Jason Calacanis, Jack Clark, Elon Musk, and Marc Andreessen are as reported in that commentary.