The Testability Line

The line between recoverable and unrecoverable work.

Lloyd Blankfein named the only question that decides whether agentic AI belongs in an institution: can you test whether it is right?

AUTHOR
Shane Schreck

PUBLISHED
May 2026

CANONICAL URL
https://publications.aleeth.com/the-testability-line/

Publication Record

Publication	Essay
Published	May 2026
Author	Shane Schreck
Canonical URL	https://publications.aleeth.com/the-testability-line/

Lloyd Blankfein named the only question that decides whether agentic AI belongs in an institution. The industry has spent two years avoiding it.

Most autonomous systems don't fail because they lack capability. They fail because no one can prove they were ever under control.

This week, on the a16z show, Lloyd Blankfein was asked about AI. He didn't reach for superintelligence. He didn't reach for job losses. He reached for something quieter and more useful. The problem with AI, he said, isn't that it's smarter than us and going to turn us into pets. It's that we don't have the ability to test whether it's right.

That sounds modest. It's the whole ballgame.

TWO KINDS OF WORK

On one side of the line, mistakes are recoverable. A bad draft. A wrong summary. A misrouted ticket. The cost is a redo.

On the other side, mistakes are not recoverable. A trade. A settlement. An attestation. The cost is the firm.

The AI industry has spent two years selling autonomy as if every task lives on the first side of that line. Most of the money lives on the second.

FORTY-FIVE MINUTES

On August 1, 2012, Knight Capital pushed a software update before the opening bell. Seven of eight servers took the new code. The eighth still carried a dormant feature from 2003. A reused flag bit woke it up.

Forty-five minutes. Four million trades across 154 stocks. Four hundred and forty million dollars, gone. One of the largest market makers in the country did not exist as an independent company by the end of the year.

No one handed the keys to a superintelligence. They handed them to an automated system nobody could test in real time. By the time they found the off switch, there was nothing left to switch off.

THE LUXURY OF A SHELF

Compare that to a story making the rounds this month. Starbucks scrapped its AI inventory system after roughly nine months of miscounting cartons and confusing one kind of milk for another.

That was the good outcome. Counting milk is recoverable. When the system got it wrong, someone caught it, because the real cartons were sitting right there to check against. The ground truth was always one shelf away, and there was a manual method to fall back to.

Most of the places we are now pointing agents have no shelf to check against, and nothing to fall back to.

THE NON-ANSWER

Ask most AI vendors how they handle this, and you get three sentences.

- "We have guardrails."
- "We follow responsible AI principles."
- "Trust us."

That is not governance. That is marketing.

A guardrail is a hope. A principle is an intention. Neither one produces what an institution actually needs when the regulator calls, when the counterparty asks, when the board wants to know what happened: evidence.

WHAT A REAL TEST IS

A test is not a vibe. A test produces proof that a third party can check without trusting the people who built the system. Three properties, or it doesn't count.

- Independent. The verifier never has to take the issuer's word.
- Durable. The proof outlives the incident, the audit, and the founder.
- Specific. It maps to the exact ways autonomous systems break, not a generic checklist.

SOC 2 did this for cloud security. PCI-DSS did this for payments. Each one turned "trust us" into "here is the certificate, verify it yourself." For autonomous systems in the era of sovereign AI, that role is empty.

THE ARCHITECTURE FOR THE EMPTY SPACE

We built for it. Institutional Control Architecture is a certification standard for autonomous systems. Seven control layers any system must establish to be certifiable. Seven failure modes that map how agents actually break. Cryptographically signed cards any third party can verify without trusting the issuer. A live certification platform, an in-environment sensor, a browser sentry, and a public registry.

It answers Blankfein's question directly. Can you test whether it's right? With ICA, the answer is a signed artifact, not a sentence in a pitch deck.

THE MARKET ALREADY KNOWS

McKinsey just sized this at five hundred to six hundred billion dollars by 2030. Thirty to forty percent of AI spending will sit under sovereignty requirements that demand exactly this kind of proof.

The autonomy pitch was AI 1.0. Hand over the keys and let it run. The AI 2.0 question is the one Blankfein asked out loud. Can you test whether it's right?

If yes, deploy. If not, don't.

The line between recoverable and unrecoverable work is the most important line in AI right now. Blankfein drew it. The industry has been refusing to.

It already has a name.

ALEETH

AI has met its match. Truth.™