

● THE INDEPENDENT THIRD EYE FOR AI

The independence the **accord left unsolved**

The White House and every major AI company just made independent external audit a pillar of AI safety. The accord doesn't say who does the auditing — and the companies being audited are choosing their own.

Roltrader Consultancy Group · 6 min read

On 29 September 2026, President Trump and the leaders of the major AI companies — Anthropic, OpenAI, Google, Meta, xAI and Nvidia — signed the White House Accord on Superintelligence. It is voluntary, and Trump called it "morally" binding. Read past the politics and one thing stands out, because it is the same thing enterprises have always relied on to trust anything important: at the centre of the accord sits *an independent external auditor or evaluator*, tasked with assessing whether each company's controls are operating as intended. The single most powerful group in technology has just agreed that AI cannot be trusted on the maker's own word. That is not a small concession. It is the whole argument for independent assurance, conceded at the top.

Bloomberg read it as a shift in the government's safety posture toward independent audit. That is right. But the accord contains a gap it does not

resolve — and it is precisely the gap that decides whether "independent audit" means anything at all.

AT A GLANCE

- 1 The accord makes an *independent external auditor* a named pillar of AI safety — the industry conceding its own word isn't enough.
- 2 It does not stipulate *who* carries out those evaluations — and the companies being audited are engaging their own.
- 3 An auditor chosen and paid by the party it audits is not independent in the sense that matters — it is the doer picking its own referee.
- 4 The expectation won't stay at the frontier labs. Every enterprise running these systems inherits the same board-level duty to prove they operate as intended.

The word doing all the work is "independent"

An accord that mandates an audit but not the independence of the auditor has mandated the easy half. The hard half — the half that makes the word mean something — is who holds the pen. The accord is explicit that it does not stipulate who carries out the third-party evaluations, and in practice the companies take responsibility for arranging their own. An auditor selected, scoped and paid by the organisation it is auditing can be diligent, expert and entirely well-intentioned, and still not be independent in the only sense that protects anyone: it has a stake in the answer. Independence is not a competence. It is the absence of a reason to prefer one result over another.

An accord that mandates the audit but not the independence of the auditor has mandated the easy half.

The doer can't pick its own referee

This is not a new problem, and enterprises solved it long ago in every domain that matters. We don't let the team that writes the accounts appoint the auditor who signs them off. We don't let the contractor certify its own building. The reason is structural, not moral: the party that produces the work — or that stands to benefit from a clean verdict — cannot be the one that decides the work is clean, because it cannot stand outside its own interest. An AI grading its own output has this problem in its purest form. A company grading its own AI, through an auditor it chose, has the same problem one step removed. The accord names the seat. It leaves open the one thing that makes the seat worth filling: that whoever sits in it answers to no one with a stake in the result.

From the frontier labs to your business

It would be easy to file this under frontier-lab politics and move on. That would be a mistake. The expectation set at the top does not stay at the top. If the makers of these systems must submit to independent, board-reviewed assessment of whether their AI operates as intended, then every enterprise that *deploys* that AI inherits the same question from its own board, its auditors and its regulators — only pointed at its own estate. When an agent is given authority to act inside your finance system, or an AI has delivered a piece of your transformation, "our vendor says it's fine" is not an answer a board can take to the people it is accountable to. The frontier labs will be audited by whoever they appoint. The enterprise needs an answer that is independent of the vendor whose system is under review.

What real independence looks like

Independent assurance of AI is not mysterious, and it does not require access to the model's inner workings. It asks the same plain questions of AI-delivered or AI-executed work that we have always asked of human work, answered

from the source rather than the producer's say-so: does every claim trace to real, captured evidence? Do the figures reconcile to the system of record? Did the agent act within its authority, and can that be shown? The checks are deterministic — rules and recomputation against your own data — not another model's opinion, because an AI checking an AI can hallucinate the check. They are done on your own infrastructure, so nothing leaves. And they are carried out by a party that neither built the AI nor delivered the work, and so has no reason to return a particular verdict. That last property is the one the accord left open, and it is the only one that can't be bought from the vendor being reviewed.

Where we sit. The accord is a good development — it makes explicit, at the highest level, that AI must answer to an independent check. Roltrader was built to be exactly that check: independent of any vendor, deterministic rather than probabilistic, verifying against the source rather than the producer's confidence, and run on your own data so nothing leaves. The industry has now agreed the referee is necessary. We are the part it hasn't solved — the independence.

We are the only independent governance-and-validation platform that referees enterprise transformation — human- or AI-delivered — without ever delivering it ourselves.