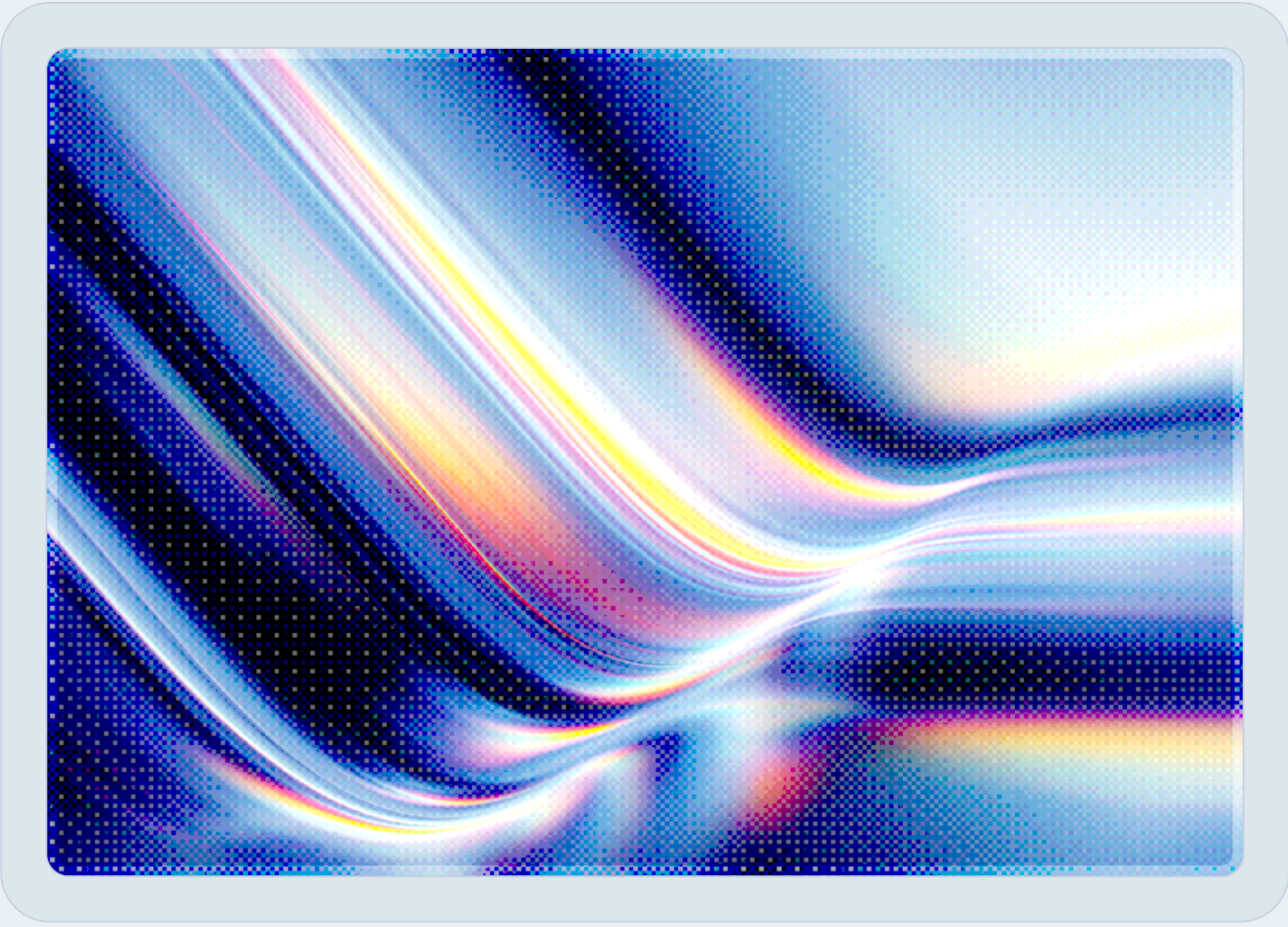
CSAI Foundation | Cloud Security Alliance

Google's FARO Proposal: A Bid to Define US AI Governance

A FINRA-Style Frontier AI Regulator Draws Rare Industry Consensus and Sharp Capture Concerns

2026-07-23

AI-assisted Rapid Research



CSAI | CSA cloud security alliance®

Key Takeaways

Google published a policy white paper, "A Pragmatic Approach to AI Governance in America," in late June 2026, proposing a two-track US regulatory model that would create a Frontier AI Regulatory Organization (FARO) — an industry-funded, federally overseen self-regulatory body modeled on FINRA — to set safety, security, incident-reporting, and transparency standards for the most capable AI systems, while leaving everyday AI applications to existing sector laws [1][2]. Google DeepMind CEO Demis Hassabis amplified the proposal on July 14, 2026, describing a body that would initially receive frontier models on a voluntary basis up to 30 days before public release and eventually require that review as a condition of US deployment [3]. The proposal has drawn unusually broad public praise across a normally adversarial industry — including from OpenAI's Sam Altman, Microsoft CEO Satya Nadella, Anthropic co-founder Jack Clark, and even Elon Musk — and reporting indicates the Trump administration, including Treasury Secretary Scott Bessent, is actively evaluating a version of the framework [4][5]. That consensus has not silenced skeptics: critics point to regulatory capture risk reminiscent of long-standing complaints about FINRA itself, an unresolved definition of "frontier AI" that could exclude systems that ought to be covered, and a compliance burden that established labs can absorb far more easily than startups or open-source developers [6][2]. The proposal also arrives without resolving — and arguably sidesteps — the deeper federal-versus-state preemption fight that has stalled in Congress since March 2026, leaving security and compliance teams to plan for a bifurcated landscape in which frontier-model oversight may eventually run through a Washington-blessed industry body while a fragmented patchwork of state AI laws continues to govern everything else [7][8].

Background

Google released "A Pragmatic Approach to AI Governance in America" around June 24, 2026 — the underlying PDF itself carries an earlier June 1, 2026 date, consistent with the paper having been finalized ahead of its public rollout [11] — framing the debate over AI regulation as a false choice between heavy-handed statutory control and an unregulated free-for-all [1][9]. The paper's central proposal splits AI oversight into two tiers. For frontier AI — the small set of the most capable, most compute-intensive models — Google proposes FARO, an independent organization with government oversight rather than a direct federal agency, funded by industry but structured similarly to established self-regulatory organizations (SROs) such as the Financial Industry Regulatory Authority (FINRA), the North American

Electric Reliability Corporation, and state bar associations [2][10]. For the much larger universe of everyday AI applications, Google argues existing law should be adapted rather than replaced, pointing to child safety, copyright, workforce transitions, privacy, and data-center energy demand as areas where targeted updates to current statutes are preferable to new AI-specific mandates [1][10].

FARO's design draws explicitly on precedent: a body governed by industry funding and staffed with independent technical experts, operating "under governmental agency oversight" in the way FINRA answers to the SEC or the American Medical Association's self-regulatory functions sit alongside federal health authorities [2][5]. Google's paper describes FARO as complementary to, rather than a replacement for, existing pre-release review activity already underway at agencies including the National Security Agency [11]. The white paper does not commit to a precise capability threshold for what counts as "frontier," an omission that at least one analyst has flagged as the proposal's most consequential gap [2].

Hassabis's July 14, 2026 remarks, reported by CNBC, added operational detail that the original white paper left open: frontier labs would initially share models with the standards body on a voluntary basis, as much as 30 days ahead of public release, with that review becoming a mandatory precondition for US deployment as the framework matures [3]. He argued the body would need "substantial" funding — likely industry-sourced — to attract technical talent and to fund the compute required for large-scale model testing [3]. Reporting since then indicates the proposal has moved beyond industry commentary into active executive-branch consideration: Fortune reports that Bloomberg identified Treasury Secretary Scott Bessent as having contributed to its development and described review by White House Chief of Staff Susie Wiles, alongside discussion of a potential Securities and Exchange Commission oversight role echoing the FINRA analogy directly [4].

The proposal's reception has been notable less for its content than for who has embraced it. Sam Altman called it "thoughtful"; Musk, who has repeatedly criticized Google's competitive position in frontier AI, described it as "a thoughtful framework overall and certainly a good starting point for discussions"; and Nadella and Block CEO Jack Dorsey offered public support [5][4]. Anthropic's Jack Clark called the framework "excellent," writing on X that "at this point, everyone at the frontier of AI agrees that third-parties should test out AI systems and use these to develop standards to feed into policy" [15]. This note reads that degree of cross-industry agreement among companies that compete aggressively on nearly every other dimension as itself a signal worth registering: a shared preference for an industry-shaped standards body over either statutory mandates or continued regulatory silence.

Security Analysis

For security and compliance teams, FARO's most consequential feature is not whether it happens — that remains genuinely uncertain — but the shape of the obligations it would impose if it does. The proposal envisions FARO setting scientific benchmarks for frontier capabilities with explicit emphasis on cybersecurity and chemical, biological, radiological, and nuclear (CBRN) risk, verifying independent audits of developer safety practices, and requiring incident reporting and transparency disclosures [2] [10]. That maps closely onto capabilities organizations should already be building toward under frameworks like the NIST AI Risk Management Framework or CSA's own AI Controls Matrix: documented model risk assessments, red-team evaluation of dual-use capability, incident response procedures specific to AI systems, and audit trails demonstrating pre-release testing. Organizations operating at or near the frontier capability threshold should treat FARO's proposed audit-and-incident-reporting regime as directionally likely regardless of whether this specific structure survives the legislative and executive-branch process, since the underlying substance — third-party testing feeding into policy, per Clark's framing — commands support among several of the most prominent voices in US frontier AI, including Altman, Nadella, Clark, and Musk.

The unresolved capability threshold is a genuine security-planning problem, not just a definitional nitpick. Critics have noted that "frontier AI" as a regulatory category is still murky enough that it could exclude systems that pose real risk while capturing others that do not, and Gartner analyst Nader Henein has argued flatly that "self-regulation is not viable" given that even well-resourced vendors lack the independent capacity for rigorous self-oversight [2][4]. Deep learning researcher Yoshua Bengio has pressed a related point: any credible version of this framework needs a defined, binding timeline for the shift from voluntary to mandatory participation, backed by independent auditing rather than developer self-attestation [4]. Until that threshold and enforcement mechanism are settled, organizations building on or fine-tuning frontier-adjacent models face ambiguity about which regulatory tier — a future FARO regime, existing sector law, or state-level AI statutes — will ultimately govern their deployment.

The proposal also surfaces a structural asymmetry that security leaders evaluating vendor risk should weigh directly. Google, OpenAI, and Anthropic already maintain the legal teams, security organizations, government relationships, and technical infrastructure a certification-style regime would require; a smaller lab, a well-resourced startup building on an open-weight model, or an open-source developer would face a materially steeper compliance climb for the same obligations [2]. That dynamic is precisely the regulatory-capture concern critics have raised by analogy to FINRA itself — Senator Elizabeth Warren has repeatedly argued FINRA structurally favors brokerages over the investors it is meant to protect, and Fortune reports that outside analysis has found FINRA failed to flag high-risk brokers despite clear

grounds for doing so [4]. A frontier AI analogue built on the same SRO template inherits the same risk: standards shaped by the entities best positioned to meet them cheaply, potentially entrenching incumbent labs' advantage regardless of the framework's stated safety rationale.

Finally, FARO's frontier-only scope leaves the underlying state-federal preemption conflict unresolved, and that conflict is where most enterprises' near-term compliance exposure actually sits. The White House's March 2026 National Policy Framework urged Congress to preempt the growing body of state AI law, but that push has stalled — Congress declined to include broad preemption in both the One Big Beautiful Bill Act and the National Defense Authorization Act, and a December 2025 executive order directing DOJ to challenge state AI laws in court had, as of the April 2026 reporting cited here, yet to produce a filed complaint [7][8]. With 145 state AI laws enacted across 38 states in 2025 alone, and frameworks like California's SB 53, Colorado's SB 205, Texas's TRAIGA, and Illinois's frontier-developer obligations already in force or advancing, a FARO-style federal frontier regime would layer on top of — not replace — that state patchwork for the overwhelming majority of AI systems that fall outside the frontier tier [8]. Enterprises should not read progress on FARO as progress toward regulatory simplification for their broader AI portfolio.

Recommendations

Immediate Actions

Security and compliance teams at organizations developing or deploying models that could plausibly be classified as frontier-capability should begin documenting the practices FARO's proposal presumes exist: pre-release risk assessments covering cybersecurity and CBRN-relevant capability, red-team evaluation results, and incident response procedures scoped specifically to AI system failures rather than general IT incidents. Organizations should separately confirm their current obligations under enacted state AI laws — including California SB 53, Colorado SB 205, Texas TRAIGA, and Illinois's frontier-developer provisions — since none of those obligations are contingent on FARO's outcome and several already carry active compliance deadlines [8].

Short-Term Mitigations

Enterprises with a stake in how frontier oversight develops should track the proposal's legislative and executive-branch path rather than treating it as settled, given reporting that places it under active White House review with a still-undefined role for the SEC [4]. Where organizations are approached to participate in voluntary pre-release model sharing of the kind Hassabis described, security teams should

ensure any such disclosure is governed by clear data handling and confidentiality terms before committing, since the mechanics of a 30-day voluntary review window have not yet been formalized [3]. Aligning existing AI governance documentation to a recognized baseline — the NIST AI RMF or CSA's AI Controls Matrix — gives organizations a defensible starting posture that would satisfy a FARO-style audit regime as well as most current state requirements, reducing the cost of whichever framework ultimately prevails.

Strategic Considerations

Organizations should plan for a durable two-tier compliance landscape rather than a near-term resolution: a possible federally sanctioned frontier oversight body running alongside a persistent, unharmonized set of state AI statutes governing everyday deployment. Multinational organizations should also weigh the international dimension of Hassabis's framing — an explicit bid for the US to define global frontier AI standards ahead of the EU AI Act's conformity assessment regime and China's own AI governance architecture — since a US-anchored FARO benchmark may not automatically satisfy obligations in other jurisdictions and could require parallel compliance tracks. Finally, organizations should factor regulatory capture risk into vendor and partner due diligence: if a FINRA-style frontier body materializes, its standards will likely be shaped disproportionately by the labs with the resources to participate in shaping them, and smaller AI vendors in an enterprise's supply chain may face compliance costs that larger incumbents do not, a dynamic worth surfacing in third-party risk assessments now rather than after a framework is finalized.

CSA Resource Alignment

FARO's core premise — that frontier AI oversight should run through an industry body operating in the space federal safety institutions have been receding from — connects directly to CSA's analysis in NIST Drops 'Safety': What the AI Consortium Rebrand Signals [13]. That research note examined the same institutional trend from the government side: NIST's 2026 renaming of its AI Safety Institute Consortium and the earlier rebranding of the US AI Safety Institute itself, both of which shifted federal AI governance language away from safety and toward innovation speed and standards competition. Read together, the two developments describe a single dynamic — federal AI safety infrastructure narrowing its stated mission at close to the same moment industry proposes to build a parallel, privately funded body to do safety-adjacent standard-setting instead. CSA's note recommended that organizations not rely on federal oversight frameworks alone and instead strengthen independent vendor due diligence and

continuous AI risk monitoring; that guidance applies with equal force to a still-unformed FARO, whose enforcement mechanism, funding structure, and mandatory-compliance timeline all remain open questions.

The voluntary, time-bound pre-release review Hassabis described — frontier labs sharing models with a standards body up to 30 days ahead of public release — is also the direct subject of CSA's US Voluntary AI Pre-Release Testing: Enterprise Governance Implications [16]. That white paper analyzes the federal government's existing shift toward voluntary pre-release testing of frontier models, covering the same 30-day access window and the same five-lab footprint — Google DeepMind, Microsoft, xAI, OpenAI, and Anthropic — that Hassabis's FARO proposal envisions folding into a more formal standards body. Because that voluntary federal arrangement is already operating rather than merely proposed, it gives security teams a closer real-world analogue for what a FARO-administered review would look like in practice than the FARO white paper's own text provides.

The unresolved federal-state tension underlying FARO's frontier-only scope is the direct subject of two companion CSA research notes: AI Preemption Battleground: Federal Framework vs. State Regulation and State AI Laws Take Hold as Federal Preemption Stalls [12][8]. Both examine the same stalled preemption effort discussed above and reach a consistent conclusion: enterprises cannot defer AI compliance planning while waiting for federal action to harmonize the landscape. That conclusion holds even if FARO advances, since its frontier tier addresses only the narrow slice of AI systems capable enough to qualify, leaving the state statutes those notes catalog — including the 145 laws enacted across 38 states in 2025 that the second note documents — governing the rest of the AI systems most enterprises actually deploy [8].

Finally, CSA's AI Controls Matrix (AICM) v1.1 [14] provides the practical control baseline organizations can use today to prepare for FARO-style obligations without waiting for the proposal's legal status to resolve. AICM's domains covering AI system lifecycle management, security testing, and incident response map closely to the audit, benchmarking, and incident-reporting functions Google's paper assigns to FARO, giving security teams a way to build toward likely future requirements using a control framework that is already published, versioned, and independent of any single regulatory outcome.

References

[1] Google. "Read our white paper on a pragmatic approach to AI governance in America." Google, June 2026.

[2] Lance Eliot. "Diving Headfirst Into The Google Newly Released 'AI Governance In America' Framework." Forbes, July 7, 2026.

[3] CNBC. "Google DeepMind chief Demis Hassabis calls for U.S. to spearhead AI standards body." CNBC, July 14, 2026.

[4] Fortune. "Google DeepMind CEO's proposal for a 'FINRA for AI' gains momentum — but is it any good?." Fortune, July 21, 2026.

[5] Digital Watch Observatory. "Google proposes a balanced approach to AI governance in the US." DiploFoundation, June 2026.

[6] The Register. "Google wants AI regulation, but on its own terms." The Register, June 26, 2026.

[7] Morgan Lewis. "White House AI Framework Puts Federal Preemption at the Center of the Debate." Morgan Lewis, March 2026.

[8] Cloud Security Alliance. "State AI Laws Take Hold as Federal Preemption Stalls." CSA AI Safety Initiative, April 4, 2026.

[9] InsideAIPolicy. "Google proposes self-regulatory body as complement to current NSA pre-release review process." Inside AI Policy, June 2026.

[10] AIFrontPage. "Amid Frontier AI Control Debate, Google's AI Governance Plan Pitches a New U.S. Regulator." AI Front Page, July 2026.

[11] Google. "A Pragmatic Approach to AI Governance in America." Google, June 1, 2026.

[12] Cloud Security Alliance. "AI Preemption Battleground: Federal Framework vs. State Regulation." CSA AI Safety Initiative, April 13, 2026.

[13] Cloud Security Alliance. "NIST Drops 'Safety': What the AI Consortium Rebrand Signals." CSA AI Safety Initiative, May 31, 2026.

[14] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2026.

[15] Jack Clark. "Post on X." X (formerly Twitter), July 2026.

[16] Cloud Security Alliance. "US Voluntary AI Pre-Release Testing: Enterprise Governance Implications." CSA AI Safety Initiative, June 20, 2026.