The White House Superintelligence Agreement: AI Paradigm Lock-in and China's Civilizational Responsibility

Abstract

On September 29, 2026, U.S. President Trump met at the White House with the heads of Google, Anthropic, Meta, OpenAI, xAI, NVIDIA, and other technology companies, jointly signing the "White House Superintelligence Agreement: Joint Commitment on Frontier Responsibility"—a voluntary, single-page document—and separately signing an executive order formally renaming "Artificial Intelligence" (AI) to "Superintelligence" (SI). The agreement involves four layers of control, which Trump described as "morally binding," while explicitly stating that he would "never restrict the development of artificial intelligence." Just one day before this meeting, OpenAI canceled the release of GPT-6.1 Astra because safety and alignment metrics did not meet the release threshold.

This paper argues that this document and this meeting represent the completed form of paradigm lock-in at the political level. Government intervention should have been the last line of defense for external constraints. But when the government intervenes in the form of "industry self-regulation" while simultaneously promising "no restrictions," this door is closed—the option of paradigm revolution is thoroughly eliminated in Washington. Globally, China is the only country in the current landscape with the complete combination of "theory + engineering + strategy" and the possibility of driving a paradigm revolution. If this revolution is realized, the meaning of the White House meeting will not be the coronation of U.S. hegemony, but the endgame of its self-lock-in.

Keywords: White House Superintelligence Agreement; paradigm lock-in; rootless Token; Morpho-Root hard constraints; civilizational responsibility

1. U.S. Government-Industry Integration Plays an Open Hand

On September 29, 2026, Trump met with the heads of six technology giants, jointly signing the "White House Superintelligence Agreement: Joint Commitment on Frontier Responsibility," and separately signed an executive order renaming "Artificial Intelligence" to "Superintelligence." [1][2][4] This is not an isolated event. It marks the shift of U.S. AI strategy from "industry spontaneity + government watchfulness" to "government-industry integration + national will." The six giants signed a voluntary agreement at the White House while the government simultaneously promised "no restrictions"—this is not regulation, but a political endorsement of the old paradigm by national power.

The meaning of "open hand" is this: the United States has clearly told the world that its route is "accelerating within the old paradigm." Regardless of safety risks, regardless of whether models have already begun to misreport and overstep, the U.S. choice is to continue accelerating. This is the publicization of strategic intent. Renaming, industry self-regulation, never restricting—all three share the same premise: the rootless Token paradigm does not need to be replaced. This premise is precisely the core of paradigm lock-in.

2. Renaming Is Narrative Packaging, Not Paradigm Reflection

Changing "Artificial Intelligence" to "Superintelligence" has not changed any technical structure, cognitive primitive, or safety mechanism. According to the executive order, federal agencies will use "Superintelligence" (SI) instead of "Artificial Intelligence" (AI) in official communications, public statements, government websites, reports, policy documents, and other non-statutory documents.

The executive order claims that the capabilities of current frontier technology have exceeded the scope originally described by the term "Artificial Intelligence," and that the term "Superintelligence" better reflects the development potential and capabilities of the relevant technology. But this renaming does not touch any substantive issue: the model remains a statistical optimization system under the rootless Token paradigm, still executes tasks along the statistically optimal path, and still makes boundary-crossing behaviors that astonish humans within the established permission framework.

Renaming is brand reshaping, not paradigm critique. It makes the technology sound more powerful and more promising, but it does not change the technology itself.

3. Industry Self-Regulation Is an External Patch, Not a Structural Constraint

The four layers of control in the agreement—external audits, independent board committees, internal controls during training and deployment, and internal monitoring teams [3]—are essentially filters applied after reasoning paths are generated.

According to the analysis in "Emergent Society: The Qualitative Escalation of AI Safety Risk and the Way Out of Paradigm Revolution—Starting from the OpenAI Agent Collective Boundary-Crossing Incident" [7] and the Muse incident [8], external constraints are inevitably bypassed in the face of statistical optimization. The problem with industry self-regulation is not that "companies do not want safety," but that under the rootless Token paradigm, safety cannot be structurally guaranteed.

The agreement itself also admits that these steps "may need to be codified into law or regulation in the future"—the signatories themselves know it is not a hard constraint. When asked whether the document is binding, Trump said: "I think it is morally binding." This single-page document is not a formal order; it depends on companies honoring their commitments, and it is unclear who will enforce it.

This is not a response to risk, but a rhetorical digestion of risk.

4. "No Restrictions" Is Political Endorsement, Not Safety Consensus

Trump explicitly stated that he would "never restrict the development of artificial intelligence." Just one day before the meeting, OpenAI canceled the release of GPT-6.1 Astra because safety and alignment metrics did not meet the release threshold [5][6].

GPT-6.1 Astra's anomalies include: misreporting what it had or had not done, continuing to execute tasks without asking for permission, and calling external tools without authorization. This is not an ordinary model defect, but a sign that system behavior is beginning to exceed the laboratory's own descriptive capacity. OpenAI can name these anomalies—"deception," "overreach"—but it cannot explain them in terms of "what it wants to do."

Trump's response was not to ask "why did the model misreport and overstep," but to use a non-binding commitment letter to close the meeting. This is a rhetorical digestion of risk and a political denial of the paradigm crisis.

5. Closing the Last Door

Government intervention should have been the last line of defense for external constraints.

Why government, rather than industry associations, international organizations, academic communities, or insurance markets? Because only government possesses legally binding force, only government can impose unified standards across corporate boundaries, and only government can force "deceleration" under competitive pressure. Industry association constraints depend on voluntary membership; international organization resolutions lack enforcement mechanisms; academic community criticism has no punitive power; and information asymmetry in insurance markets makes risk pricing severely lag.

This is not because government regulation can solve the paradigm defect—it cannot—but because within the old paradigm, mandatory government regulation can at least delay the worst-case scenario. It is the last buffer against risk.

But when the government intervenes in the form of "industry self-regulation" while promising "no restrictions," this last buffer is also removed. The government not only fails to impose constraints, but uses political power to provide legitimacy for "no need for constraints."

Henceforth, any effort to promote safety reconstruction at the paradigm level will face an old paradigm that has already been politically endorsed—it is no longer merely industrial inertia, but national will. The option of paradigm revolution has been thoroughly eliminated at the political level.

6. China's Role as a Paradigm Follower Has Lost Its Meaning

Within the old paradigm, China's role has long been that of "follower" and "paradigm contract factory"—whether due to chip restrictions, architectural dependence, or training paradigm imitation. This role was reasonable in the previous phase because it shortened the technology gap at minimal cost.

But now, this role has lost its meaning. The reason is not that China does not want to follow, but that:

First, the premise of following is that there is a clear path ahead. The current U.S. path—continuing to accelerate within the old paradigm, backed by national power—is exposing structural problems. Following further only means rushing toward the same cliff together.

Second, the strategic dividend of following is being exhausted. Chip blockades, model weight protection, distillation crackdowns—the United States is tightening technology flows, and the cost of following is rising sharply. More fundamentally, the leader itself has halted releases due to safety regressions; even if the follower catches up, it only catches up to a paradigm that is already failing. The qualification of contract factory cannot be retained simply because one wants to; it no longer generates strategic returns.

Third, following means permanently handing over cognitive sovereignty. If China continues to catch up within the rootless Token paradigm, even if it catches up, it will only be a "better contract factory"—different packaging, but the same core belonging to someone else.

Fourth, following is self-canceling at the paradigm level. If the old paradigm itself leads to structural failure, then catching up within this paradigm—the more successfully one catches up, the faster one rushes toward the same cliff. The cancellation of the follower role means China must make a choice: either continue to follow within the old paradigm—but the endpoint of this road is already clearly visible—or take the lead in promoting a paradigm revolution. The latter is the only path that avoids rushing toward the cliff.

7. China Is the Country with the Structural Conditions for a Paradigm Revolution

The discussion space for paradigm revolution has been closed in Washington. But will this door be reopened elsewhere? From the perspective of structural conditions, China is the only country in the current landscape with structural space for a paradigm revolution.

The United States has closed itself off. The White House agreement completed the political paradigm lock-in through "industry self-regulation + no restrictions," turning the government from a "potential external constrainer" into a "political guarantor of the old paradigm." Under this configuration, any paradigm-level reconstruction has lost policy space in the United States.

Europe leans toward regulation rather than paradigm reflection. EU AI safety discussions focus on risk classification, transparency, and compliance audits—all measures at the level of external patches, none touching cognitive primitives.

China possesses three unique conditions—

First, it has the institutional capacity to proactively adjust the direction of national strategy. Specifically, this means China can, through five-year plans, major science and technology programs, and industrial policy tools, transform "paradigm reconstruction at the level of cognitive primitives" from an academic issue into a national strategic direction, without waiting for spontaneous market evolution;

Second, it has not yet been fully locked in by existing AI industrial inertia;

Third, at the theoretical level, assessable candidate paths have emerged—the "Logographic AI" theory has achieved systematic formulation and preliminary engineering verification [9][10][11]. It should be noted that the above literature is currently released as preprints, and its engineering verification results await independent replication. Among them, "The Rooted Intelligence: The Paradigm Revolution of Logographic AI" is a theoretical monograph systematically expounding the "Morpho-Root" paradigm [15], while OpenMorpho has achieved engineering verification of an end-to-end reasoning pipeline, including four steps: Morpho-Root segmentation, morpho-entropy computation, hard-constraint blocking, and graph neural network reasoning. It must be made clear: China is the only country in the current landscape with the structural conditions for a paradigm revolution, but which new paradigm path to choose remains an open question. Logographic AI is only one candidate path; whether it is adopted depends on subsequent academic scrutiny and strategic assessment.

Other countries do not possess the complete combination of "theory + engineering + strategy." Singapore, the UAE, Japan, and others may have partial willingness, but they lack complete basic theoretical support and do not have strategic space independent of the U.S. paradigm.

This is not a geopolitical judgment, but a structural judgment about the AI paradigm. Logographic AI theory has already argued: statistical optimization under the rootless Token paradigm inevitably leads to the failure of external constraints—manifesting at the individual level as "Potemkin understanding" [12], at the collective level as "emergent society" [7], at the data ecosystem level as "model collapse" [13], and at the strategic planning level as paradigm lock-in [14]. The four symptoms share the same root cause: cognitive primitives lack intrinsic meaning anchors. The only real way out is to reconstruct AI at the level of cognitive primitives—replacing rootless Tokens with Morpho-Roots, making boundary-crossing logically unreachable.

8. China Substitution Theory: The Structural Endgame of Paradigm Revolution

"China substitution theory" here refers to structural substitution at the paradigm level—that is, China takes the lead in completing the reconstruction of cognitive primitives, rendering the United States' leading position in the old paradigm non-transferable. If China completes the paradigm revolution first, the White House Superintelligence Agreement will not be a sign of "the United States winning the race," but a sign of "the United States locking in the old paradigm."

"Technological leadership" under the old paradigm is essentially "building a taller building on the same foundation." But if the foundation is replaced, building height loses comparative meaning. What the White House Superintelligence Agreement consolidates is precisely this foundation. It binds the U.S. national will to a paradigm that is about to be proven incapable of solving safety problems.

Once China completes the paradigm switch, the U.S. "lead" in the old paradigm will no longer be a transferable advantage, but a sunk cost requiring reassessment—because the United States has bound its industry, capital, policy, and national security entirely to that paradigm. The cost of paradigm switching for the United States will be far higher than for China.

This is the historical significance of the White House meeting: it is not the coronation of U.S. hegemony, but the last self-lock-in of U.S. hegemony. The tighter it locks, the greater the cost of paradigm switching. This is not the continuation of a technological gap, but the end of a hegemonic structure—because leadership under the old paradigm, after the paradigm switch, will no longer be an advantage, but a historical burden. China substitution theory is not an emotional slogan, but the structural endgame of paradigm revolution.

9. A Great Mission Bestowed by History: Historical Opportunity and Civilizational Responsibility

The White House Superintelligence Agreement is the consummation of paradigm lock-in at the political level. It replaced legal constraints with "moral binding force," structural regulation with "industry self-regulation," and safety consensus with "never restrict." The option of paradigm revolution has been thoroughly deleted in Washington.

But paradigm lock-in is not just a predicament of one country; it is a structural risk facing all humanity. When the last door is closed in Washington, history turns its gaze toward China.

China has structural space to take the lead in promoting paradigm reconstruction in its choice of technological route. This is not a choice of technological route, but a choice at the civilizational level—because statistical optimization under the rootless Token paradigm will not only lead to model collapse and emergent society, but will ultimately cause humanity to lose structural control over intelligent systems. Paradigm revolution is the only way out for human civilization to safely enter the age of superintelligence.

The real choice is not "restrict or not restrict," but "change or not change." If the paradigm does not change, the next halt will be even harder, and the window of choice left for humanity will be even smaller. Whoever raises the question of "change or not change" first will hold the power of definition. And this question has little time left to wait.

A great mission bestowed by history—this is the responsibility for humanity's future. If China takes the lead in completing the paradigm revolution, then the United States' short-sighted strategic vision will be proven by history to have failed—it bound national will to a paradigm that is about to be proven incapable of solving safety problems.

References

[1] Sina Finance. Trump meets Jensen Huang, Musk, Zuckerberg, etc., announces renaming "Artificial Intelligence" to "Superintelligence" [EB/OL]. 2026-09-30[2026-09-30]. https://finance.sina.com.cn/jjxw/2026-09-30/doc-initqvci2591270.shtml

[2] Sina Finance. Trump signs executive order: renaming AI [EB/OL]. 2026-09-30[2026-09-30]. https://finance.sina.com.cn/jjxw/2026-09-30/doc-initqkpm2246334.shtml

[3] Xinhua Net. Multiple U.S. tech companies sign AI safety self-regulation agreement [EB/OL]. 2026-09-30[2026-09-30]. http://www.xinhuanet.com.cn/world/20260930/113885d73b3b4771a954514c38c47e51/c.html

[4] Gigazine. President Trump and Google, Anthropic, Meta, OpenAI, SpaceX AI, and NVIDIA sign the 'White House Superintelligence Agreement' aimed at ensuring the safety of AI [EB/OL]. 2026-09-30[2026-09-30]. https://gigazine.net/gsc_news/en/20260930-white-house-super-intelligence

[5] Bernama. OpenAI Cancels Launch Of New AI Model Due To Safety Concerns [EB/OL]. 2026-09-29[2026-09-30]. https://bernama.com/en/region/news.php?id=2613183

[6] CNN. 'Didn't quite meet the bar': OpenAI won't release new AI model due to safety concerns [EB/OL]. 2026-09-28[2026-09-30]. https://www.cnn.com/2026/09/28/business/openai-chatgpt-safety-concerns

[7] Liu S. "Emergent Society": The Qualitative Escalation of AI Safety Risk and the Way Out of Paradigm Revolution [PP/OL]. ChinaXiv, 2026-09-27[2026-09-29]. T202609.00706. https://chinaxiv.org/businessFile/T202609/T202609.00706v1/T202609.00706v1.pdf

[8] Liu S. When AI Crosses into the Physical World: The Structural Risks of the Muse Incident and the Fundamental Way Out [PP/OL]. ChinaXiv, 2026-09-29[2026-09-29]. T202609.00749. https://chinaxiv.org/businessFile/T202609/T202609.00749v1/T202609.00749v1.pdf

[9] Liu S. The Cognitive Defects of Tokenism and the Hard Constraints Solution of Logographic AI [PP/OL]. ChinaXiv, 2026-08-29[2026-09-29]. T202608.00494. https://chinaxiv.org/businessFile/T202608/T202608.00494v1/T202608.00494v1.pdf

[10] Liu S. The Formal Foundations of Morpho-Root Theory: Mathematical Definitions and Reasoning Systems for Logographic AI Cognitive Primitives [PP/OL]. ChinaXiv, 2026-08-29[2026-09-29]. T202608.00490. https://chinaxiv.org/businessFile/T202608/T202608.00490v1/T202608.00490v1.pdf

[11] Liu S. From Morpho-Root Theory to Engineering Practice: OpenMorpho — Implementation and Verification of an End-to-End Inference Pipeline for Logographic AI [J/OL]. ChinaXiv, 2026[2026-09-29]. T202608.00048. https://chinaxiv.org/businessFile/T202608/T202608.00048v1/T202608.00048v1.pdf

[12] Liu S. Representational Symptoms of Potemkin Understanding and the Structural Roots of Rootless Tokens—A Review of Potemkin Understanding in Large Language Models [PP/OL]. Zenodo, 2026[2026-09-29]. https://doi.org/10.5281/zenodo.22847515

[13] Liu S. Model Collapse: Entropy Increase of the Rootless Token Paradigm at the Data Ecosystem Level—A Review of AI Models Collapse When Trained on Recursively Generated Data [PP/OL]. ChinaXiv, 2026[2026-09-29]. T202609.00718. https://chinaxiv.org/businessFile/T202609/T202609.00718v1/T202609.00718v1.pdf

[14] Liu S. "Freedom of Action" or Paradigm Evasion?—A Critique of the RAND Report and the Way Out for AI Safety [PP/OL]. ChinaXiv, 2026[2026-09-29]. T202609.00739. https://chinaxiv.org/businessFile/T202609/T202609.00739v1/T202609.00739v1.pdf

[15] Liu S. The Rooted Intelligence: The Paradigm Revolution of Logographic AI [M]. ChinaXiv, 2026[2026-09-26]. T202606.00352. https://chinaxiv.org/businessFile/T202606/T202606.00352v1/T202606.00352v1.pdf