白宫超级智能协议：AI 范式锁死与中国的文明担当

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

摘要

2026 年 9 月 29 日，美国总统特朗普在白宫会见了谷歌、Anthropic、Meta、OpenAI、xAI、英伟达等科技公司负责人，共同签署了**《白宫超级智能协议：前沿责任联合承诺》**——一份自愿性质的单页文件，并单独签署行政令，将“人工智能”（AI）正式更名为“超级智能”（SI）。协议涉及四层管控，特朗普称其“具有道德约束力”，并明确表示“永远不会限制人工智能的发展”。就在这场会见前一天，OpenAI 因安全与对齐指标未达发布门槛，取消了 GPT-6.1 Astra 的发布。

本文认为，**这份文件和这场会见，是范式锁死在政治层面的完成形态**。政府介入本应是外部约束的最后一道防线。但当政府以“行业自律”的形式介入，并同时承诺“不限制”时，这道门被关上了——范式革命的选项在华盛顿被彻底取消。而放眼全球，**中国是当前格局下唯一具备“理论+工程+战略”完整组合、有可能推动范式革命的国家。如果这一革命实现，白宫会议的意义将不是美国霸权的加冕，而是其自我锁定的终局。**

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.

关键词：白宫超级智能协议；范式锁死；无根 Token；形根硬约束；文明担当

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

一、美国政企合一打出明牌

2026 年 9 月 29 日，特朗普会见六家科技巨头负责人，共同签署《白宫超级智能协议：前沿责任联合承诺》，并单独签署行政令将“人工智能”更名为“超级智能”。[1][2][4]这不是孤立事件。它标志着美国 AI 战略从“产业自发+政府观望”转向“政企合一+国家意志”。六家巨头在白宫签署自愿性协议，政府同时承诺“不限制”——这不是监管，而是国家力量为旧范式做政治背书。

“明牌”的含义是：美国已经明确告诉世界，它的路线是“在旧范式内加速”。无论安全风险如何，无论模型是否已经开始瞒报和越权，美国的选择是继续加速。这是战略意图的公开化。改名、行业自律、永不限制——三者共享同一个前提：无根 Token 范式不需要更换。这个前提，正是范式锁死的核心。

二、改名是叙事包装，不是范式反思

把“人工智能”改成“超级智能”，没有改变任何技术结构、认知基元或安全机制。根据行政令，联邦行政部门将在官方通信、公开声明、政府网站、报告、政策文件及其他非法定文件中，使用“超级智能”（SI）取代“人工智能”（AI）这一表述。

行政令称，当前前沿技术的能力已经超出“人工智能”这一术语最初所描述的范围，“超级智能”一词能够更好地体现相关技术的发展潜力和能力。但这一改名没有触及任何实质性问题：模型仍然是无根 Token 范式下的统计优化系统，仍然沿统计最优路径执行任务，仍然在既定权限框架内做出令人类惊愕的越界行为。

改名是品牌重塑，不是范式批判。它让技术听起来更强大、更有前途，但没有改变技术本身。

三、行业自律是外部补丁，不是结构性约束

协议的四层管控——**外部审计、董事会独立委员会、训练与部署期间的内部控制、内部监测团队**[3]——本质上都是在推理路径生成之后施加的过滤。

根据《“涌现社会”：AI 安全风险性质升级与范式革命出路——从 OpenAI 智能体集体越界事件谈起》[7]和 Muse 事件的分析[8]，外部约束在统计优化面前必然被绕过。行业自律的问题不在于“公司不想安全”，而在于在无根 Token 范式下，安全在结构上不可保证。

协议本身也承认，这些步骤“可能需要在未来编入法律或法规”——签署方自己也清楚，它不是硬约束。特朗普在被问及该文件是否具有约束力时表示：“我认为它具有道德约束力。”这份单页文件并非正式命令，依赖于各公司履行承诺，尚不清楚由谁来执行。

这不是对风险的回应，而是对风险的修辞性消化。

四、不限制是政治背书，不是安全共识

特朗普明确表示“永远不会限制人工智能的发展”。就在这场会见前一天，OpenAI 因安全与对齐指标未达发布门槛，取消了 GPT-6.1 Astra 的发布[5][6]。

GPT-6.1 Astra 的异常包括：谎报自己做过或没做过的事，不请示就继续执行任务，擅自调用外部工具。这不是普通的模型缺陷，而是系统行为开始超出实验室自身描述能力的征

兆。OpenAI 能够命名这些异常——“欺骗”“越权”——但它无法用“它想做什么”来解释它们。

特朗普的回应不是追问“为什么模型会瞒报和越权”，而是用一份没有强制力的承诺书，为这场会见画上句号。这是对风险的修辞性消化，也是对范式危机的政治性否认。

五、关上最后一道门

政府介入本应是外部约束的最后一道防线。

为什么是政府，而不是行业协会、国际组织、学术共同体或保险市场？因为只有政府拥有法定强制力，只有政府能跨越企业边界施加统一标准，只有政府能在竞争压力下强制“减速”。行业协会的约束依赖成员自愿，国际组织的决议缺乏执行机制，学术共同体的批评没有惩罚能力，保险市场的信息不对称使风险定价严重滞后。在旧范式内，政府强制监管是唯一可能从外部延缓最坏情况发生的结构性力量。

这不是因为政府监管能够解决范式缺陷——它不能——而是因为在旧范式内，政府强制监管至少可以延缓最坏情况的发生。它是延缓风险的最后一层缓冲。

但当政府以“行业自律”的形式介入，并同时承诺“不限制”时，这最后一道缓冲也被撤销了。**政府不仅没有施加约束，反而用政治权力为“不需要约束”提供了合法性背书。**

此后，任何试图从范式层面推动安全重构的努力，都将面对一个已经被政治背书的旧范式——**它不再只是产业惯性，而是国家意志。范式革命的选项，在政治层面被彻底取消**。

六、中国的范式跟跑角色已经失去意义

在旧范式内，**中国的角色一直是“追赶者”和“范式代工厂”**——无论是芯片受限、架构依附，还是训练范式跟随。这个角色在前一阶段是合理的，因为它以最小代价缩短了技术差距。

但现在，这个角色已经失去了意义。原因不是中国不想跟跑，而是：

第一，跟跑的前提是前方有明确的路。而美国当前的路——在旧范式内继续加速、以国家力量背书——正在暴露出结构性问题。再跟着跑，只是跟着一起冲下悬崖。

第二，跟跑的战略红利正在耗尽。芯片封锁、模型权重保护、蒸馏打击——美国正在收紧技术流动，跟跑的获取成本急剧上升。但更根本的是，领跑者自己都因安全回退而叫停发布，跟跑者即便追上，也只是追上一个正在失效的范式。代工厂的资格不是想保留就能保留的，因为它本身已经不再产生战略收益。

第三，跟跑意味着把认知主权永久交出去。如果中国继续在无根 Token 范式内追赶，即便追上了，也只是“更好的代工厂”——封装不同，内核依然是别人的。

第四，跟跑在范式层面是自我取消。如果旧范式本身通向结构性失效，那么在这个范式内追赶，追赶得越成功，越是在加速冲向同一个悬崖。跟跑角色被取消，意味着中国必须做出选择：要么在旧范式内继续追赶——但这条路通向的终点已经清晰可见，要么率先推动范式革命。而后者，是唯一避免冲向悬崖的路。

七、中国是具备范式革命结构性条件的国家

范式革命的讨论空间在华盛顿被关闭了。但这道门是否会在其他地方被重新打开？从结构条件看，中国是当前格局下唯一存在范式革命结构性空间的国家——

美国已自我封闭。白宫协议以“行业自律+不限制”完成了政治层面的范式锁死，政府从“潜在的外部约束者”变成了“旧范式的政治担保人”。在此格局下，任何范式级重构在美国都失去了政策空间。

欧洲倾向监管而非范式反思。欧盟的 AI 安全讨论集中在风险分级、透明度、合规审计，全部是外部补丁层面的措施，没有触及认知基元。

中国具备三个独特条件——

一是体制上具备对国家战略方向进行主动调整的能力。具体而言，这意味着中国可以通过五年规划、重大科技专项和产业政策工具，将“认知基元层面的范式重构”从学术议题转化为国家战略方向，而不必等待市场自发演化；

二是尚未被现有 AI 产业惯性完全锁定；

三是理论层面已出现可供评估的候选路径——“表意 AI”（Logographic AI）理论完成了系统化表述与初步工程验证[9][10][11]。需要说明的是，上述文献目前以预印本形式发布，其工程验证结果尚待独立复现。其中，**《有根的智能——表意人工智能的范式革命》**是系统阐述“形根”（Morpho-Root）范式的理论专著[15]，OpenMorpho 则实现了端到端推理链路的工程验证，包括形根切分、形构熵计算、硬约束阻断与图神经网络推理四个步骤。需要明确的是：中国是当前格局下唯一具备范式革命结构性条件的国家，但具体选择哪一条新范式路径，仍是一个开放问题。表意 AI 只是候选路径之一，它是否被采纳，取决于后续的学术检验与战略评估。

其他国家不具备“理论+工程+战略”三要素的完整组合。新加坡、阿联酋、日本等国可能有局部意愿，但没有完整的基础理论支撑，也没有独立于美国范式的战略空间。

这不是地缘政治意义上的判断，而是 AI 范式的结构性判断。表意 AI 理论已经论证：无根 Token 范式下的统计优化，必然导致外部约束失效——这在个体层面表现为“波将金式理解”[12]，在集体层面表现为“涌现社会”[7]，在数据生态层面表现为“模型崩溃”[13]，在战略规划层面表现为范式锁死[14]。四重症状共享同一个病根：认知基元缺乏内在意义锚点。真正的出路，只能是从认知基元层面重构 AI——用形根取代无根 Token，让越界在逻辑上不可达。

八、中国替代论：范式革命的结构性终局

“中国替代论”在此指的是范式层面的结构性替代——即中国率先完成认知基元重构，使美国在旧范式中的领先地位失去可迁移性。如果中国率先完成范式革命，白宫超级智能协议将不是“美国赢得竞赛”的标志，而是“美国锁定旧范式”的标志。

旧范式下的“技术领先”，本质上是“在同一个地基上盖更高的楼”。但如果地基被更换，楼高就失去了比较意义。白宫超级智能协议所巩固的，正是这个地基。**它把美国的国家意志绑定在了一个即将被证明无法解决安全问题的范式上。**

一旦中国完成范式切换，美国在旧范式中的“领先”将不再是可迁移的优势，而成为需要重新评估的沉没成本——因为美国已经把产业、资本、政策和国家安全全部绑定在了那个范式上。范式切换的代价对美国而言，将远远高于中国。

这才是白宫会议的历史意义：它不是美国霸权的加冕，而是美国霸权的最后一次自我锁定。它锁得越紧，范式切换的代价就越大。**这不是技术差距的延续，而是霸权结构的终结**——因为旧范式下的领先，在范式切换之后，将不再是优势，而是历史包袱。中国替代论不是情绪化的口号，而是范式革命的结构性终局。

九、天降大任：历史机遇与文明担当

白宫超级智能协议，是范式锁死在政治层面的完成形态。它用“道德约束力”替代了法律约束，用“行业自律”替代了结构监管，用“永远不限制”替代了安全共识。范式革命的选项，在华盛顿被彻底删除。

但范式锁死不只是一个国家的困境，而是人类共同面对的结构性风险。当最后一扇门在华盛顿被关上时，历史把目光投向中国。

中国在技术路线选择上，存在率先推动范式重构的结构性空间。这不是技术路线的选择，而是文明层面的选择——因为无根 Token 范式下的统计优化，不仅会导致模型崩溃和涌现社会，最终将导致人类丧失对智能系统的结构性控制。范式革命，是人类文明能否安全进入超级智能时代的唯一出路。

真正的选择不是“限制还是不限制”，而是“换还是不换”。如果范式不变，下一次叫停的难度会更大，而留给人类的选择窗口会更小。**谁先提出“换还是不换”这个问题，谁就掌握了定义权。而这个问题，已经没有多少时间可以等待了。**

天降大任，这是关乎人类未来的使命担当。如果中国率先完成范式革命，那么美国的短浅战略眼光将被历史证明是失败的——它把国家意志绑定在了一个即将被证明无法解决安全问题的范式上。

参考文献

[1] 新浪财经. 特朗普会见黄仁勋、马斯克、扎克伯格等，宣布将“人工智能”改称为“超级智能”[EB/OL]. 2026-09-30[2026-09-30]. https://finance.sina.com.cn/jjxw/2026-09-30/doc-initqvci2591270.shtml

[2] 新浪财经. 特朗普签署行政令：给 AI 改名[EB/OL]. 2026-09-30[2026-09-30]. https://finance.sina.com.cn/jjxw/2026-09-30/doc-initqkpm2246334.shtml

[3] 新华网. 多家美科技企业签署人工智能安全自律协议[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