



校准赤字

为什么中国卓越的治理能力越来越难以及时感知现实，并在压力演变为危机前完成调整

中国拥有人类历史上最强大的协调系统之一。然而，使其能够进行大规模动员的同一治理架构，也越来越削弱了其在复杂性条件下进行准确感知、适应性反馈与系统性校准的能力。

比约恩·肯尼斯·霍尔姆斯特伦

2026年5月

知识共享署名-相同方式共享 4.0 国际许可协议

国家报告·中国

<https://bjorkennethholmstrom.org/reports/china-calibration-deficit>

执行摘要

2020年1月，中国官方对一种新型病原体进行了基因组测序，并在几天内向全球分享。几周之内，他们封锁了一座拥有1100万人口的城市。几个月内，通过大规模检测、接触者追踪、定向封锁以及其他任何政府都无法企及的行政动员能力，中国在其境内基本抑制了病毒。从流行病学标准来看，这是国家协调行动的一次卓越成就。

两年后，曾以非凡能力应对疫情初期的同一系统，在短短一周内彻底放弃了其整个政策框架。“动态清零”政策——通过三年的大规模隔离、封锁公寓楼以及对每一步行动的二维码监控来执行——在没有任何过渡计划、公共卫生准备或对即将发生的情况进行坦诚说明的情况下，一夜之间被逆转。随后引发的感染浪潮在几个月内导致估计100万至150万人死亡，如果医疗系统能提前获得准备时间，其中许多人本可以幸免于难。这次失败并非源于能力。中国拥有医院、药物和物流基础设施。失败在于时机：由于循序渐进的修正需要承认之前的错误，一项早已被证据证明无法维持的政策无法得到逐步调整。

这两次事件并非矛盾。它们是同一种底层状况的两种表现。这就是**“校准赤字” (Calibration Deficit)**：不是能力的短缺，而是一种结构性的无能，导致国家的现实模型无法在最重要的领域以最快的速度与现实本身保持同步。

中国真正做对的地方。 在进行诊断之前，必须诚实地承认其成就。在四十年的时间里，中国实现了一个治理系统在人类历史上从未以如此规模和速度实现的成就：将14亿人口的社会从贫困转变为接近发达世界的生活水平。八亿人脱贫。拥有世界最大的高铁网络。令西方相形见绌的数字支付基础设施。产业政策的协调使中国在多个领域成为主导制造商。这些是真实的成就，如果一个治理框架不能解释威权协调在特定条件下能达到的效果，那它就不是一个严肃的框架。发展奇迹是诊断的基础，而非对其的反驳。

“运动—过头—突发修正”循环。 中国的治理具有特征性的节奏。中央设定优先事项，行政机器以彻底且惊人的速度动员起来。但这种动员会产生自身的扭曲：职业生涯取决于合规表现的地方官员往往过度执行，并隐瞒障碍。现实情况与上报信息之间的差距先是缓慢扩大，随后迅速拉开，直到跨越阈值。修正随之而来——但它是“逆转”而非“调整”的形式出现的，因为渐进式调整需要承认那些从未被承认的积弊。房地产扩张演变为债务危机；科技行业增长演变为监管风暴；“动态清零”演变为突然放弃；反腐运动导致官僚机构普遍出现寒蝉效应。每一次都遵循同样的弧线。这不是政策的不一致，而是一个拥有超强执行能力但反馈架构存在系统性缺陷的治理架构的结构性输出。

“校准赤字”的定义。 中国能以民主系统只能望其项背的果断和规模执行政策。它可以集成——党对政治合法性的垄断消除了分散的合法性冲突。它可以动员——中央可以以联邦制国家无法企及的速度和完整性，使省、市和机构行动者步调一致。它无法可靠做到的是更新其行动所依据的模型。

“校准赤字”是指国家的现实模型与现实本身之间的差距，这一差距之所以扩大，是因为在差距影响最重大的领域，处理反馈的架构受到了系统性的削弱。系统在执行其认为正确的事情时表现出色，问题在于它所相信的事实变得越来越不准确——而本应纠正这一点的架构，恰恰也是系统生存逻辑要求其压制的架构。

巴西与中国的镜面案例。 巴西陷入困境是因为权力过于分散而无法协调；中国受到限制是因为权力过于集中而无法自纠。两者共同定义了本系列研究所描绘的“协调与适应性”光谱的两个极端。发展奇迹证明了拥有有效反馈的集中权威可以产生非凡的发展成果；而“运动一过头一突发修正”循环则证明了反馈受损的集中权威所产生的成本，与其失准的持续时间成正比。

结构性机制。 七个互锁的机制产生并再现了“校准赤字”。**“晋升锦标赛”在可见的短期目标上创造了近乎完美的契合，但在难以衡量或政治敏感的现实上造成了脱节——地方官员成了治理的“演技大师”而非真正的考题解决者。地方政府融资平台（LGFV）债务引擎将职业驱动的乐观主义转化为表外负债，在土地出让收入崩溃和债务到期同时发生前隐秘堆积。认识论反馈崩溃——由于权力的巩固逐步消除了内部多元化——缩减了向上呈报信息的视角范围，而此时人口和债务压力恰恰最需要准确的感知能力。算法现实扭曲场恶化了治理分析所依赖的数据，因为接受了净化数据训练的AI系统学会了观察一个被管理过的现实，而非现实本身。人口硬件崩溃在校准能力最脆弱的时刻增加了一场加速测试——独生子女政策的后果现在是结构性的，无法通过设计消除。台湾变量将最重要的资源分配决策锁闭在反馈架构之外，成为一种永久性的治理扭曲，无法在当前的合法性框架内进行诚实的成本效益分析。而“明哲保身”**这一儒家原则确保了那些对治理失败拥有最准确私人认知的人，最不可能向上反映。这些机制共同构成了一个系统性的校准失败架构。

诚实的框架。 本报告基于功能性而非规范性立场。论点并非中国应效仿西方民主，而是中国自身的目标——持续的经济增长、政权稳定、技术领先、人口韧性、国家安全——与一个日益受损的校准架构是不兼容的。无论对这些目标的价值判断如何，“校准赤字”都威胁到了中国既定目标的实现。系统需要的不是西方意义上的民主化，而是功能性校准能力所需的特定治理改革：受保护的反馈渠道、可逆的决策结构、奖赏准确性而非仅奖赏步调一致的激励机制，以及恢复邓小平时代开发并被此后逐步压制的实验性联邦制。

真正的自适应能力需要什么。 本报告最重要的分析视角，是将处方置于中国自身的先例而非外国模式中。“摸着石头过河”的方法论、作为校准工具的经济特区、作为分布式自适应响应的乡镇企业——这些不是西方进口品。它们是中国自己的创新，通过维持反馈架构，使地方实验能够产生成功的方法，并由中央进行观察、比较和推广，从而创造了发展奇迹。论点不是“要像西方一样”，而是“恢复曾成就发展奇迹的校准能力，并将其扩展到目前被排除在外的领域”。

控制保存本能。 抵制这些改革的免疫系统不是被动的制度惯性，而是一种主动的、嵌入在架构中的自我保存机制，它准确地将校准改革识别为对当前系统运行逻辑的威胁。受保护的反馈渠道会创造出其评估可能与中央战略方向相矛盾的制度行为者。可逆的决策结构会将“承认错误”内置于政策程序中。激励机制改革会创造出敢于呈报不悦真相的官员群体。将“稳定”与“控制”混为一

谈的学说——即将分散的权力视为动荡的前兆——确保了每一项改革都被免疫系统处理为生存威胁，而非治理改进。这种免疫系统在其逻辑内部并非不理性，但它是自我消解的：对改革的压制确保了校准赤字继续积累阈值事件，最终迫使改革在危机而非深思熟虑的条件下发生。

具体的第一步。 在当前的架构下，功能性校准能力所需的改革受到结构性限制。在这些约束下可行的第一步是“实验治理协议”：由国务院正式指定一系列省市作为受保护的实验空间，授权地方官员在其中实施新型治理方法，而不必承担失败对职业生涯的影响，评估标准基于所获得的经验教训而非既定结果。这不是外国建议，而是以正式制度保护的方式恢复定义了改革开放时代并创造了发展奇迹的实践。此外，“绩效准确性审计”提供了一种补充机制：在现有的考核框架内，引入对官员报告准确性的正式评估，随着时间的推移，在报告准确性与职业晋升之间建立可见的相关性，从而改变目前使“粉饰报告”成为理性选择的计算方式。

系列研究的边界条件。 本报告是涵盖德、法、瑞典、印、欧盟、英、巴西、俄、美、芬兰系列研究中的第十一篇。中国在这一分类中占据特殊位置。与俄罗斯一样，这是该系列框架触及边界条件的两个案例——在这些点上，分布式权威和功能性反馈的处方无法在当前政权下转化为可行的转型架构。但中国绝非俄罗斯。俄罗斯的治理架构已经摧毁了改革赖以生存的制度底色，而中国在逐步损害反馈架构的同时，保留了非凡的制度能力。区别在于“在当前政权下受限，但可修复”，而非“在当前政权下不可能”。不同之处在于，中国的治理传统中包含了更具自适应能力的治理记忆和先例。修复是可能的，至于是否以及何时选择修复，则不是本框架所能回答的问题。

模拟与领土。 中国构建了国家曾建造过的最精密的社会模型之一。数据系统、监控基础设施、行政报告机制、计划机制——它们共同构成了一个广度极广、技术极其先进的模型。“校准赤字”就是这个模型与它所代表的真实领土之间发生的进行性脱节。晋升锦标赛过滤了进入模型的信息；地方债架构创造了模型财政表征中未捕捉到的负债；认识论反馈崩溃缩小了模型构建者的视角；“明哲保身”确保了那些私人知识与模型表征偏差最大的人最不可能揭示这种偏差。模拟系统以非凡的复杂性和力量运行，而现实领土无论模拟系统如何描述，依然按照其自身的规律存在。

结语。 发展奇迹是人类治理的重大成就之一。本报告无法回答，但现实领土最终会回答的问题是：创造了这一奇迹的治理系统，是否能创造出中国及世界所需的下一个奇迹。本系列框架提供的不是预测，而是一个可能未来的架构：中国治理系统可以拥有这一未来，因为邓小平时代的模式证明了中国系统具备这种能力，且这种能力与非凡的国家成就并存。当前系统与该架构之间的距离，衡量了权力巩固项目所付出的代价。第一步是在目前可用的领域、利用目前允许的机制，开始缩小这一距离，并以一种积累证据的方式进行，使其价值在显现时，成为开启未来的基石。

1. 校准赤字 (The Calibration Deficit)

1.1 一个“双面”国家

2020年1月，武汉报告了一组病因不明的肺炎病例。几天之内，中国官方就完成了病原体的基因组测序并与世界分享。几周之内，他们封锁了一座拥有1100万人口的城市。几个月内，通过大规模检测、接触者追踪、定向封锁，以及动员了全球任何其他政府都无法企及的行政能力，中国在其境内基本抑制了病毒。从任何流行病学标准来看，早期的应对都是一项真正的成就——比许多富裕的民主国家更快、更协调、更有效。

两年后的2022年12月，曾以非凡能力管理疫情初期的同一系统，在短短一周内彻底瓦解了其整个政策框架。“动态清零”——这一执行了三年且严厉程度不断升级的策略（通过大规模隔离、封锁公寓楼、对每一步行动进行二维码监控来强制执行）——在没有任何过渡计划、没有公共卫生准备、也没有对即将发生的情况进行坦诚说明的情况下被废除。在政策取消后的几个月里，随之而来的感染浪潮导致估计100万至150万人死亡，如果医疗系统能提前获得准备时间，其中许多人本可以幸免于难。这次失败并非源于能力。中国拥有医院、药物和物流基础设施。失败在于时机：由于循序渐进的修正需要承认之前的错误，一项早已被证据证明无法维持的政策无法得到逐步调整。

这两次事件——疫情初期的应对与动态清零的崩溃——并非矛盾。它们是同一种底层状况的两种表现。中国并不缺乏行动能力。它能以民主系统只能望其项背的果断和规模采取行动。它无法可靠做到的是更新其行动所依据的“模型”——无法通过其治理架构察觉到情况已经改变、策略已经失效、且必须在跨越阈值之前进行修正。这就是**“校准赤字”**：不是能力的短缺，而是一种结构性的无能，导致国家的现实模型无法在最重要的领域以最快的速度与现实本身保持同步。

1.2 “运动—过头—突发修正”循环

中国的治理具有一种特征性的节奏，这与本系列研究过的任何其他系统都不同。它不像瑞典那样“漂流”，压制信号直至其演变成危机；它不像德国那样“停滞”，因自身程序的摩擦而瘫痪；它也不像巴西那样在“突破”与“利益俘获”之间循环。它是“踉跄”前行——在一段坚定、协调、高速度的行动之后，隐藏成本不断堆积，随后的修正不是逐渐到来的，而是以“逆转”的形式出现。

这种循环具有可辨识的结构。首先设定一个中央优先事项：消灭新冠、扩张房地产以驱动增长、培育国内技术冠军、消除党内腐败。一旦优先级确立，行政机器便以令人惊叹的彻底性动员起来。资源涌入，官员竞争，地方政府执行。这种动员是实实在在的，而非表演性的。中国将国家力量凝聚

在特定目标背后的能力，是人类历史上最强大的治理工具之一。

但动员会产生自身的扭曲。地方官员的职业生涯取决于对中央优先事项的服从，因此他们有动力过度执行，却极度排斥汇报障碍。负债超过可持续水平的房地产开发商会继续建设，因为停下来就标志着失败。观察到动态清零正在造成社会经济破坏的官员会非常谨慎地汇报这些破坏（如果他们真的汇报的话），因为该政策是中央的重中之重，而质疑中央优先事项并非一种有利于职业发展的行为。现实情况与上报信息之间的差距先是缓慢扩大，随后迅速拉开，直到跨越阈值——例如一个足以威胁金融系统的开发商倒闭、一场足以引发街头抗议的残酷封锁，或是计划中被战场现实瞬间推翻的军事假设。

在阈值点，修正到来了——但它是“逆转”而非“调整”的形式出现的，因为渐进式调整需要承认那些此前从未被承认的积弊。房地产政策、动态清零、科技行业整顿、使官僚机构丧失主动权的反腐运动：每一项都遵循同样的弧线。运动在初期取得了真实的成果，随后动员过头，成本在暗处堆积。修正到来得太晚且太突然。系统重新稳定，并为下一场运动做准备。

这并非领导力糟糕或意识形态僵化的症状（尽管两者都会放大这种现象）。这是一个拥有超强执行能力但反馈架构存在系统性缺陷的治理架构的结构性输出。一个能动员十亿人实现中央目标，却无法可靠察觉到该目标需要修正的系统，是一个行动成本累积速度超过行动修正速度的系统。这就是“校准赤字”最精确的表述。

1.3 “校准赤字”的定义

本系列之前的报告诊断出的故障模式在不同程度上都是“能力”的失败。德国无法执行是因为行政机器支离破碎且缓慢；法国无法整合是因为其治理架构排斥持久政策所需的分布式合法性；瑞典无法及时察觉是因为其共识文化在干扰信号到达决策者之前就将其过滤掉；印度无法同步是因为其规模超出了体制架构的协调能力；英国无法交付是因为决策权与执行力处于系统的不同层级；巴西无法积累是因为掠夺性的政治架构在每一个突破产生复利效应前就将其消耗殆尽。

中国的问题本质不同，而非程度不同。中国可以执行，其基础设施项目按时完成且规模令其他政府艳羡。它可以整合——党对政治合法性的垄断消除了撕裂法国或美国治理的分布式合法性冲突。它可以在特定领域和条件下进行感知：早期的疫情应对展示了大多数政府无法企及的快速信号处理能力。它可以同步——动员架构允许中央以联邦制国家无法接近的速度和完整性使地方与机构步调一致。它可以凝聚——党提供了一定程度的意识形态和组织凝聚力，这是像欧盟这样的超国家机构只能仰望的。

中国所缺乏的——其治理架构在结构上无法提供的是：**对其行动与这些行动所产生的现实之间进行持续校准。**“校准赤字”是指国家的现实模型与现实本身之间的差距，这一差距之所以扩大，是因为在差距影响最重大的领域，处理反馈的架构受到了系统性的削弱。该系统在执行其认为正确的事

情时设计得极为精妙。但在一个复杂、动态、日益互联的系统中，它在根据现实更新其认知模型方面的能力存在结构性缺陷。

这不同于在俄罗斯诊断出的信息压制。俄罗斯压制信息是一种生存策略——“权力纵向”刻意摧毁反馈架构以防止竞争性权威中心的出现。中国的校准赤字则更微妙，且在某些方面更危险。中国的领导层并不希望压制负面信息。党明确推崇“实事求是”。问题不在于意图，而在于架构：在这个治理系统中，各级官员的职业激励机制系统性地奖赏准确的好消息而过滤准确的坏消息；中央权威如此绝对，以至于对中央战略方向的异议需要一种在系统激励下极度稀缺的政治勇气；随着权力的集中（习近平的权力巩固加速了这一进程），本应在问题达到阈值前登记积弊的反馈回路被逐步收窄。

1.4 中国与所有其他案例的区别

上述观点需要面对一个显而易见的质疑：中国经济在四十年的时间里以历史上前所未有的速度增长；其基础设施是世界上最先进的之一；它让8亿人脱贫，这是有记录的历史上在最短时间内最大程度减轻人类痛苦的成就。如果中国存在校准赤字，这些成就是如何实现的？

答案是：校准赤字是特定领域的，并随着权力集中而日益严重。产生中国发展奇迹的邓小平时代系统，其反馈架构比当前系统更有效。“摸着石头过河”的方法论是一种真正的自适应治理实践：地方实验、中央观察、选择性推广、对差异的包容。经济特区是制度多样化的刻意实验。20世纪80年代的乡镇企业代表了一种真正的多中心经济适应。当时的校准赤字虽然存在，但受到了约束。

所改变的——以及将当前系统与产生经济奇迹的发展型国家区分开来的是：允许反馈运行的领域被逐步缩小。权力巩固加速了自1989年后就已经显现的趋势：现实可以反驳中央战略方向的空间正在稳步萎缩。在改革早期，经济反馈相对自由。尝试失败经济模式的地方官员被允许失败；找到成功模式的官员被效仿。市场提供的反馈，政治系统在不威胁其主要合法性来源（经济增长）的情况下无法压制。这种经济反馈回路创造了发展奇迹。

但经济反馈回路会与政治反馈回路相互作用。房地产行业膨胀到不可持续的水平，部分原因是政治激励机制奖赏地方官员的土地出让收入，而不顾长期可持续性。科技行业在其全球影响力达到巅峰时被遏制，是因为政治考量凌驾于经济校准之上。一项疫情政策在流行病学证明失效后仍被维持，是因为在系统的激励架构中，承认需要修正的政治成本超过了政策持续的成本。校准赤字并非静止的。随着权力集中增加了政治考量凌驾于经济和社会反馈之上的领域，赤字会不断复合增长。

最能概括这一点的一句话是：**巴西陷入困境是因为权力过于分散而无法协调；中国受到限制是因为权力过于集中而无法自纠。**

1.5 镜鉴

中国这一案例直接挑战了本系列隐含的论点。之前的国别报告共同构建了一个案例，即分布式权威、真正的辅助性原则、有效的反馈回路和自适应治理在功能上优于权力集中、中央协调和行政动员。中国是一个反证。其治理架构在设计上几乎与本系列推荐的模式截然相反——但它产生的结果，是本系列推崇的治理模式在同等规模和速度下显然未能实现的。

面对这一挑战，诚实的做法不是否定或辩解中国的成就。这些成就是真实的。诚实的做法是精确指出在何种条件下集中权威优于分布式治理，以及在何种条件下则不然。拥有功能性反馈架构的集中权威（邓小平时代系统在经济领域所接近的状态）可以实现分布式系统需要几十年才能完成的大规模快速转型。发展奇迹就是证据。

而反馈架构受损的集中权威（当前系统随着诚实反馈空间缩小而日益接近的状态）则会产生以动态清零崩溃为代表的“运动—过头—突发修正”循环。能力依然非凡，但这种能力被部署的方向与现实日益脱节。修正即便到来，也往往太晚且代价高昂。

本报告问的问题不是中国是否应该成为西方自由民主国家。这不仅在分析上毫无用处，本系列也不会给出答案。问题在于：中国自身设定的目标——持续发展、政权稳定、技术领先、国家韧性——是否与一个日益受损的校准架构相兼容。本报告得出的答案是否定的：校准赤字威胁到了中国实现其自身目标的能力，这与任何关于这些目标应是什么的规范性判断无关。这一框架使本报告能够被那些原本可能将其视为西方意识形态说教的读者所理解，且更重要的是，这在分析上是诚实的。

1.6 核心问题

如果不强行简化其复杂性，就无法用单一诊断框架描述一个拥有14亿人口的治理系统。中国包含着多重面相：一个能在大陆上建成高铁网络的国家，和一个无法可靠告诉其领导层在一场它已宣告胜利的疫情中死了多少人的国家。一个拥有极深制度底蕴的系统，和一个拥有极大制度脆弱性的系统。一个真正改善了人类史上最多人的生活条件的政党，和一个系统性地破坏了分布式制度基础（而这些基础本可以让下一阶段的发展像上一阶段一样快速和规模化）的政党。

“校准赤字”是连接这些表面矛盾的线索。它不是对中国的全面描述，它是对中国未来几十年走势最具决定性的特定结构特征——这一特征决定了系统拥有的非凡能力，是被一种对该能力“该做什么”的日益准确的理解所引导，还是被一个日益脱离其所宣称代表的现实领土的模型所引导。

本节定义了“校准赤字”，追踪了它在“运动—过头—突发修正”循环中的表现，将其与本系列在其他地方诊断出的故障模式区分开来，正视了中国成就这一反证，并建立了分析框架：即基于功能必要性而非规范性处方。接下来的章节将诊断产生并再现校准赤字的结构机制：晋升锦标赛与大规模的委托代理失败、地方债引擎、认识论反馈崩溃及其在权力巩固下的加速、算法现实扭曲场、人口硬件约束、作为治理扭曲的台湾变量，以及早于且将长久于当前政权的“明哲保身”文化操作系统。随后，报告将描述功能性校准架构的要求——不是西方意义上的民主化，而是中国自身目标所要求的特定治理改革——并以本系列一贯的坦诚，承认当前架构下的“可能”与“不可能”，以及这种区别为何至关重要。

2. 校准赤字：结构性机制 (Structural Mechanisms)

2.1 什么是“校准能力”

任何治理系统都必须履行两种截然不同的功能，二者之间的关系决定了系统的长期生存能力。第一是**执行 (Execution)**：将决策转化为结果、动员资源以实现既定目标、将政治意志转化为机构行动的能力。第二是**校准 (Calibration)**：登记这些结果是否符合目标、决策所依据的现实模型是否依然准确、以及在“模型”与“现实领土”之间的差距演变为灾难前修订行动的能力。

中国的治理架构是人类历史上最强大的执行引擎之一，但其校准能力却受到了系统性的削弱。本节描述的机制解释了其原因——这并非个人失败或意识形态错误的结果，而是一个治理架构的结构性输出。该架构的激励设计、信息流和权力分配针对“执行”进行了优化，而这种优化在功能上与持续、诚实的“校准”是不兼容的。理解这些机制是理解任何改革之必要条件的先决条件——也解释了为什么校准架构的改革会威胁到依赖于它的执行架构的基础。

2.2 晋升锦标赛与大规模的委托代理失败

任何大国的治理架构都面临着根本性的“委托—代理”问题：中央设定目标，但外围负责执行；外围的利益——职业考量、资源约束、地方政治压力——并不会自动与中央目标对齐。每个大国都有管理这种失调的机制，中国的机制则是“晋升锦标赛”。

中国系统内各层级的官员——省长、市委书记、县长——都在一套由党的组织部门管理的绩效考核体系下竞争晋升。考核标准虽历经数十年演变，但其基本结构在“可见的、可量化的、短期目标”上创造了近乎完美的契合，而在“长期的、难以衡量的或政治敏感的现实”上造成了近乎完美的脱节。

这种结构带来的职业后果已有详尽记录。一个在其辖区内实现高 GDP 增长、完成基础设施建设并维持社会稳定的官员会获得晋升。而一个汇报其 GDP 增长是通过债务堆积实现（这些债务将成为下一任的问题）的官员，不会因为诚实而晋升；他们只有在任期内避免债务暴露才能晋升。一个看到辖区内环境恶化的官员会非常谨慎地汇报（甚至不汇报），因为驱动恶化的经济活动同时也驱动了决定其仕途的 GDP 数据。当一名官员意识到中央在特定领域的战略方向正产生有害后果时，他面临抉择：诚实汇报伤害并冒着被视为对战略方向不忠的风险，或是粉饰报告并冒着伤害不断累积直至跨越无法管理的阈值的风险。

在数十年间、数百万名官员的规模下，这种激励结构产生了一个可预测的总体结果：到达中央的信息被系统性地歪曲，偏向于中央“想听到的”，而非中央“需要知道的”。这种歪曲并非阴谋，而是数百万代理人在其运作的激励架构内做出的个体理性反应。每一项具体的粉饰报告行为在个体层面都是可辩解的，但总体效应是中央的现实模型与现实本身发生了偏离——起初缓慢，随着偏离的复合增长而迅速拉开。

这就是校准赤字核心的委托代理失败。这并非中国所独有，每个大型组织都会面临这类问题。中国案例的特殊之处在于系统的规模，以及晋升锦标赛缺乏结构性的制衡力量。在民主系统中，反对党、自由媒体、独立法院和公民社会组织提供了让“令人不悦的信息”到达决策者的替代渠道。这些渠道虽然不完善且常有扭曲，但它们确实存在。在中国的当前架构中，这些制衡力量被系统性地削弱或消除，使得晋升锦标赛成为信息向上传递的主导机制——并随之带来了该锦标赛激励结构所产生的所有扭曲。

2.3 地方政府融资平台（LGFV）债务引擎

晋升锦标赛的扭曲在地方政府的财政架构中最为明显。中国地方政府承担着基础设施投资、社会服务和经济发展的重担，而这些构成了普通公民所体验到的政府活动的主体。然而，现行财政结构将收入集中在中央，而将支出责任下放到地方——地方官员必须通过正式预算之外的手段来弥补这一缺口。

弥补这一缺口的大规模机制就是“地方政府融资平台”（LGFV）。LGFV 是地方政府通过其借款以资助基础设施、土地开发和经济活动的表外实体，这些活动产生了仕途晋升所需的 GDP 增长。由于借贷是通过 LGFV 而非正式预算进行的，它不会出现在官方债务统计中。由于偿还主要依赖于土地出让收入和未来的经济增长，其可持续性取决于地方官员有动力乐观预测、却无机制能保证的各种条件。

LGFV 系统不仅是一个财政问题，它更是一个校准问题。它是一种机制，将晋升锦标赛的激励结构——“通过展示增长来晋升，隐瞒成本”——转化为中国经济的物理和金融架构。LGFV 资助的项目往往是真实的且有价值的，但资助它们的债务也是真实的，且累积速度超过了土地出让和增长预测所能长期维持的水平。当房地产行业减速时（如 2021 年以来的结构性减速），用于偿还 LGFV 债务的收入随之崩溃，数十年表外借贷累积的债务压力便会同时爆发。

LGFV 债务引擎说明了校准赤字结构性机制中反复出现的一个原则：在当前激励架构中真实存在但不可见的成本会不断堆积，直到跨越某个无法再通过常规渠道管理的阈值。到那时，“运动一过头一突发修正”循环再次显现——这并非政策失败，而是结构性的必然，是系统性激励产生隐形负债的架构的可预测输出。

中央并非没有意识到 LGFV 问题。连续几轮的去杠杆运动、债务置换计划和监管收紧相继推出，每一轮都在降低可见风险方面取得了一定成效，但产生风险的底层激励结构却保持不变。这就是在制度层面运作的校准赤字：系统可以识别问题并动员力量应对其症状，但无法重新配置产生问题的激励架构，因为该架构同时也是系统实现其主要合法性来源（增长目标）的机制。

2.4 认识论反馈崩溃

晋升锦标赛和 LGFV 债务引擎是一个更普遍现象的具体表现：即中国政府处理自身信息的反馈架构正在逐步崩溃。这种崩溃并非新鲜事，但在习近平的权力巩固下，其加速方式与早期的信息管理实践有着质的区别。

邓小平时代的治理系统维持了一定程度的内部多元化，即使不能提供完全准确的信息，至少提供了竞争性的信息。党内派系在经济战略、产业政策和开放节奏上持有真实的分歧。这些派系间的竞争通过政治局常委会的集体领导结构进行调节，创造了异议观点到达最高决策层的渠道。这虽非民主审议，但它是以“功能性信息处理”为副产品的精英竞争。其结果是一个能够进行重大转向的治理系统——如 1978 年改革开放、1992 年南巡、2001 年加入 WTO——每一次都代表了中央现实模型在应对累积证据时的重大修订。

而权力的巩固逐步拆解了这种内部多元化。反腐运动无论其初衷如何，在结构上都消除了党内潜在的替代性战略视角的来源。决策权的集中减少了能通过异议迫使审议的独立博弈参与者。将意识形态一致性提升为晋升资格，对那些本可以通过专业领域知识引入现实校正视角的专业官员造成了筛选压力。

其结果是，系统最高层级的校准架构变窄了。这并非因为当前领导层想要坏信息——党的文件反复强调“实事求是”的重要性——而是因为让坏消息在层级晋升过程中得以幸存的结构性条件被削弱了。一名官员要向政治局常委会陈述不悦的真相，不仅需要勇气，更需要一个保护此类言论的制度环境。随着制度环境的收窄，传递真相所需的认识论勇气增加了，而愿意行使这种勇气的官员供应量则下降了。顶层的模型与底层的现实日益脱节，这并非源于单一的欺骗行为，而是源于数百万个微小的、个体理性的“管理真相”行为的累积效应。

2.5 算法现实扭曲场

认识论反馈崩溃还有第二个维度，这是当前时代所特有的，且没有历史先例：即数字威权主义与人工智能的结合。中国用二十年时间建立了一个世界上最复杂的数字监控和信息管理架构。这些系统设计的初衷是管理公众的信息环境：防止敏感内容流传、识别并压制异议、塑造政治舆论。

由此产生的校准赤字首先作用于民众。公民无法获取关于治理失败的信息，看不到关于抗议或政策灾难的报道，他们发展出的现实模型与所处的现实是部分脱节的。公众叙事与私人知识（人们在日常生活中观察到的情况）之间的差距，往往产生的是愤世嫉俗而非真正的无知。大多数中国公民明白经过整理的环境并不反映全部。信息管理系统真正阻止的是“私人知识”向“集体意识”的汇聚——即让零散的个人观察转化为能够促成政治行动的共有理解。

但更微妙且更危险的是，这种赤字也作用于治理系统本身。中国政府正日益依赖 AI 模型、行政分析系统和预测性治理工具，而这些工具是在经过信息管理架构过滤的数据上进行训练的。为了政治敏感性而被过滤的数据、反映宣传叙事而非观察现实的数据、被晋升锦标赛激励结构塑造过的数据——这就是治理架构用于自我理解的系统之训练环境。国家不仅是在管理公民的信息环境，它正逐步管理着“自己的”信息环境。在管理过的数据上训练出的模型，学会的是观察一个管理过的现实。

这就是“算法现实扭曲场”：一个自我强化的认识论陷阱。国家在数字信息管理上的投入，正在逐步降解国家自身可用的信息质量。治理系统不仅是在过滤新闻，它在毒化自己治理所需的训练数据。而数字工具的复杂性意味着这种过程是隐形的——扭曲被编码在数据中，而非显现在审查中，其后果表现为微妙的错误模型、产生微妙的错误预测、生成微妙的错误决策，其成本在扭曲模型看不见的阴影中不断堆积。

2.6 人口硬件崩溃

上述结构性机制是治理架构的特征，原则上是可以改革的。人口约束则不同：它是治理架构并未选择也无法通过设计消除的物理现实。它对校准赤字的意义在于，它将在系统反馈架构最无法诚实应对的时候，强加一场最严苛的测试。

中国人口已开始下降，劳动年龄人口自 2012 年以来持续减少。总和生育率处于极低水平，这意味着下降将在本世纪中叶加速，无论政策如何干预。老年抚养比正在上升，这将从根本上改变养老和医疗系统的财政算术。支撑了四十年奇迹的人口红利——即大规模、年轻、城镇化的劳动力——已经耗尽。

从 1980 年起实施的独生子女政策塑造了这场危机的具体形态：一代独生子女面临规模异常庞大的老年抚养负担。在 1980 年的条件下该政策有其经济逻辑，但其在人口证据早已表明需要逆转后仍维持了很久，这本身就是校准赤字的体现：一项数据早已表明需要修订的政策，因政治架构难以修订而维持，等修订到来时已太晚。

人口约束与校准赤字的交织方式极具后果。党的“绩效合法性”——即通过提供增长和改善生活来换取认同——是建立在一个已不复存在的人口结构之上的。支撑发展奇迹的治理架构是为劳动力供应扩张、快速城镇化和出口驱动增长的时代校准的。现在这些条件都在逆转。系统面临的校准挑战

不是增量适配，而是根本性的架构修订——而当前的反馈架构，由于其系统性偏向汇报好消息和推迟坏消息，在结构上无法诚实地处理这种修订。

领导层意识到人口挑战，2016年以来的鼓励生育政策证明了这一点。但经过受损反馈架构处理的意识，产生的政策响应是为“叙事需求”而非“人口现实”校准的。鼓励生育的激励措施在任何同类社会中都未能在统计学上逆转生育率下降（韩国、日本、新加坡的证据是不含糊的）。一个拥有功能性校准能力的系统本应登记这些证据并修订政策。而在证据失效后仍坚持此类激励，是一个诊断信号：系统对“需要一个解决方案”这一叙事要求的响应，超过了对“到底什么方案有效”这一证据的响应。

2.7 作为国内治理扭曲的台湾变量

每个治理系统都在分配注意力和资源，而这种分配从非中性——它反映了系统架构所嵌入的优先级层级。在中国，一个单一的领土问题已变得如此核心，以至于它扭曲了资源分配，限制了战略灵活性，并成为诚实处理国内治理挑战的结构性障碍。

台湾问题不仅是外交政策问题。它是一个如果不经修订就会产生巨大政治成本的合法性主张，它不受党控制的时间线约束，且要求不断升级的军事准备，这些资源需求直接与应对人口和经济挑战所需的福利、教育和创新投入竞争。

校准赤字在此处以特定方式运作：台湾变量在系统的合法性架构中被设定为“不可谈判”，这意味着它无法像其他资源分配决策那样接受严谨的成本效益分析。如果一名官员提议评估军事准备成本是否与该变数的概率和后果相匹配，他实际上是在问“党的内核合法性是否值得这个代价”。在当前架构下，这个问题无法被诚实地提出，这意味着它所要求的资源分配无法根据变化的情况重新校准。系统锁定在了一条军事投入的轨道上，其最优性无法被评估，因为评估本身需要一种架构在结构上禁止的反馈。

2.8 “明哲保身”——文化操作系统

本系列中的每个国家都有一个文化锚点。芬兰是 *sisu* 与“安静共识”；巴西是 *jeitinho*；俄罗斯是 *ne vysovyvaysya*。中国则是“明哲保身”。这个词出自《诗经》，它在三千年治理史中的长寿本身就具有诊断意义——证明了中国国家机器一贯奖赏谨慎、惩罚显眼。

这个概念捕捉了一种特定形式的自适应自我审查。它不同于俄罗斯式的“别出头”（那是在掠夺性国家面前的生存本能），“明哲保身”在儒家传统中带有一种积极的智慧：智者知道何时言、何时默，何时行、何时待。一个看到政策失败的贤明官员不会公开宣布这一点，而是在适当的时间、通过适

当的渠道、带着适当的恭敬默默行事。

在拥有功能性反馈架构的系统中，这种特质会产生一种有节制的沟通。但在合法渠道被逐步收窄的系统中，它导致了对治理失败最了解的人的集体沉默。看到 LGFV 债务堆积而保持沉默的官员并非懦夫，他是理性行为者，因为在当前系统中，直言会危及仕途却无法改变结果。积累在数百万官员、科学家和工程师身上，这种个体理性的谨慎产生了一个对自己状况日益盲目的治理系统——并非成员缺乏知识，而是向上传递知识成为了一种理性人不会承担的风险。

“明哲保身”早于且将长久于当前政权。这使其比前述结构性机制更具承重性。结构可以改革，但文化操作系统对干预有着极强的抵抗力。改变它不仅需要重构激励系统，还需要维持这种重构足够长的时间，直到“直言而不会受损”成为常态，从而改变理性人的计算方式。这是一个世代工程。

2.9 国际孤立带来的反馈损失

上述机制主要是对内的。但还有一个最终机制复合了所有其他因素：即中国治理系统与外部反馈渠道的逐步脱离，而外部渠道本可以部分补偿内部反馈的缺陷。

任何治理系统都嵌入在国际环境中，国际环境通过贸易关系揭示竞争力，通过学术网络交流知识。历史证明，中国奇迹的一部分源于对这些外部渠道的开放：与国际经济机构的接触、大规模派遣留学生、整合进全球供应链。这些提供了补充性的、有时能纠正国内扭曲信号的外部校准。

因此，削减这些联系——收紧学术交流、限制外国 NGO、追求“自主”技术战略以及地缘政治压力下的脱钩——不仅是外交选择，更是“校准选择”。其后果体现在治理系统可用信息的质量上。一个收窄了国内反馈架构同时又减少外部反馈暴露的系统，正变得日益依赖于内部生成的现实模型。

2.10 机制如何相互强化

校准赤字不是上述机制的总和，而是它们的**乘积**。

晋升锦标赛创造了奖赏好消息、管理坏消息的激励。LGFV 是这种激励的财政表现，将职业乐观主义转化为隐形负债。认识论崩溃是权力巩固消除内部多元化的结果。算法扭曲场则是这种崩溃的技术放大，降解了治理数据。人口崩溃在校准能力最脆弱时施加了测试。台湾变量锁定了资源分配。而“明哲保身”确保了能修正模型的私有知识保持在私有状态。最后，国际孤立移除了可能暴露这些扭曲的外部参考点。

“运动—过头—突发修正”循环是这一交互系统的动态表达。中央设定优先级，动员是真实的且令人印象深刻。执行机器是非凡的。但登记成本、边际收益递减和副作用的反馈被上述机制过滤和压制。模型与领土的差距先是隐秘地扩大，随后跨越阈值，修正以迟到且高昂代价的方式到来。

这就是结构诊断层面的校准赤字。中国能以任何其他系统都无法企及的力量进行执行，但它尚未能维持一种诚实的反馈架构，使这种力量能被对“该做什么”的准确理解所引导——而它之所以不能，是因为其执行如此之好的原因，在结构上与削弱校准的原因纠缠在一起。

3. 功能性校准的要求 (What Functional Calibration Would Require)

3.1 原则：功能性必要而非规范性处方

前一节描述了一个校准能力受到系统性且自我强化的损害的治理架构。本节将探讨恢复这种能力需要什么——这不是出于民主理想主义的考量，而是对中国自身既定目标所要求的严谨分析。

这一框架非常重要，值得明确阐述：本报告并不主张中国应采用西方的自由民主制度。这种论点在分析上并无益处，也不是本系列研究的初衷。本系列诊断了三大洲民主国家的治理失败，其中一些最严重的失败——美国的“整合赤字”、英国的“控制与交付错位”、巴西的“积累赤字”——都发生在形式上民主的系统中。在这一系列研究形成的观点中，民主既不是良好治理的充分条件，也不是决定一个治理系统能否适应复杂性的首要变量。

本节提出的论点更为具体且紧迫：中国自身的目标——持续的经济发展、政权稳定、技术领先、人口韧性、国家安全——与一个日益受损的校准架构是不兼容的。实现这些目标中的每一个都需要关于“什么有效、什么无效”的准确信息；需要有在持续执行的成本超过变革成本之前修订战略；需要一个能够以挑战演进的速度更新其现实模型的治理系统。校准赤字威胁到了这些目标，这与对这些目标本身的价值判断无关。

这种框架使得本节的分析能够被那些原本可能将其视为西方处方的读者所接受：所描述的改革之所以被提出，不是为了让中国更像西方，而是因为它们是实现中国自身目标所必需的治理架构功能要求。

3.2 邓小平时代系统的正确之处

本报告最重要的分析视角——也是将其与标准的西方对华治理批评区分开来的一点——是将处方置于中国自身的先例而非外国模式中。支撑了发展奇迹的校准架构并非西方进口品，它是由中国领导层务实地开发出来的创新。这代领导层从“大跃进”和“文革”的灾难性校准失败中吸取了教训，明白一个无法登记自身失败的治理系统将会大规模地制造失败。

“摸着石头过河”的方法论是一个真正的自适应治理原则。它授权了无需中央许可的地方实验，创造了可观察的成果差异，使中央能够识别优劣，并建立了一种务实的修订规范，明确将政策修订与政治羞辱脱钩。经济特区不仅是经济实验，它们更是校准工具——是用于在将全国经济投入开放之

前，生成关于“开放经济会产生什么结果”的信息的装置。乡镇企业并非由中央设计，它们产生于地方的主动性，并被允许按其自身的条件成败，从而产生了一个任何中央计划机构都无法复制的分布式校准过程。

这种校准能力的制度表现是集体领导制——即通过政治局常委会进行决策的规范。在这种规范下，没有哪位领导人的观点是自动权威的，技术型高级官员之间的观点竞争提供了功能性的信息处理过程。该系统虽然不是民主的，但它拥有功能性的反馈架构，使得在证据充分积累时能够进行重大的航向修正：如1978年的市场改革、1992年的加速推进以及2001年的加入WTO。

因此，恢复校准能力的论点并不是在争论外国模式。它是在主张恢复中国治理系统自身开发、随后又逐步放弃的东西。处方不是“变得像丹麦一样”，而是“恢复曾成就发展奇迹的校准能力，并将其扩展到目前被排除在外的领域”。

3.3 受保护的反馈渠道

功能性校准架构最根本的要求是存在特定渠道，使得关于决策后果的准确信息能够到达决策者手中，而不被晋升锦标赛的激励结构所过滤。这并不要求建立自由媒体或政治反对派合法化，它要求刻意创造出诚实评估受到保护而非惩罚的制度空间。

中国治理传统对此已有先例。历史上的“谏议”制度——即指定官员有权利且有义务向皇帝提供批判性评估——是内置于帝国治理系统中的校准机制。当代党和国家的纪律检查委员会履行了这一功能的某种版本，尽管它们主要关注个人腐败而非系统的政策评估。国家发改委进行情景分析，尽管其产出受到其审查的各部委政治约束的影响。

受保护反馈渠道所要求的改革，不是创建新机构，而是刻意保护现有机构免受目前扭曲其产出的激励结构的影响。具体而言：负责预测、评估和检查功能的官员的晋升标准，需要与执行层级中的行政官员的标准挂钩脱钩。一个职业生涯取决于其所评估政策之成败的官员，无法对这些政策进行诚实评估。一个晋升取决于辖区GDP增长的官员，无法诚实汇报实现该增长的方法已造成了将在下一任期爆发的负债。

其机制是制度分离：设立负责诚实评估的机构，其成员因准确性而非步调一致获得奖励，其职业轨迹独立于受评部门，且其产出必须提交给决策者，而无需经过受评方的预先过滤。这是其他治理系统中独立央行、议会预算办公室和审计机构背后的原则。中国拥有这些机构的版本，要求是使这些版本从其评估性能的行政官僚机构中获得真正的独立。

政治障碍是巨大的。晋升锦标赛的激励结构也是党管理行政机器行为的主要机制。将评估功能与锦标赛脱钩，面临着创造出利益与中央优先级背离的制度参与者的风险——而当前的架构正是为了防止这一点而设计的。这就是根本性的紧张关系：校准能力所要求的改革，恰恰是执行架构所抵制的

改革。

3.4 可逆的决策结构

功能性校准架构的第二个要求是：在持续执行的成本超过变革成本之前修订决策的制度能力。“运动—过头—突发修正”循环不仅仅是信息不灵的产物，它也是决策结构的产物，这些结构使得增量修订在政治上的成本高于维持现状直到跨越阈值。

在当前的架构中，重大战略方向在党的最高层确定，然后作为约束性优先事项向下传递。一名建议修订战略方向的官员，实际上是在暗示领导层错了——架构将此视为一种会产生职业后果的不忠。其结果是，修订往往以“逆转”而非“调整”的形式出现，因为只有当积累的成本变得无可否认时，修订的政治成本才会降至维持现状的成本之下。

可逆决策结构将通过在原始决策中内置“修订触发点”，而不是将修订视为初始错误的证据，来改变这种动态。其机制是“试点先行”的政策设计：重大的战略倡议先在特定的地理或行业领域实施，并预先设定明确的评估标准，然后再推向全国。评估标准应是客观且公开的——不应受那些职业生涯取决于倡议成败的官员的解释。如果试点未达到标准，全国性的推广就会推迟，且这不构成政治失败；试点正按设计运行，生成了它旨在生成的信息。

中国对此也有历史先例。经济特区模式正是如此：一个带有明确对比标准的有界实验，允许中央在承诺国家政策之前评估结果。改革时代生产性地使用了这一机制数十年。它的逐步被放弃——转向全国同步启动的大规模运动，而非顺序试点——是校准赤字最清晰的结构表现之一，也是最容易改革的领域之一。

可逆决策结构的政治可行性取决于在文化上将“修订”与“失败”脱钩。这是一项重大的文化转变，但并非不可能。邓小平时代的规范——“实事求是”——明确支持将修订视为对证据的忠诚，而非对错误的承认。恢复这一规范，并构建通过试点设计和公开评估标准使其操作化的制度结构，将增加修订的政治空间，而无需进行当前架构下代价极其高昂的领导层认错。

3.5 激励改革：奖赏准确性而非仅奖赏步调一致

第三个要求在结构上最具挑战性，也是最重要的：改革目前的晋升标准，这些标准目前奖赏与中央优先级的一致，而惩罚对令人不悦现实的诚实汇报。

晋升锦标赛不是一个可以随手改革的设计特征。它是党管理一个拥有数千万官员、横跨大陆规模国家的行政机器的主要机制。在没有替代方案的情况下拆除它将创造行政真空——在这个系统中，官员将不知道哪些行为受奖、哪些受罚，且目前补充正式晋升标准的非正式网络和个人关系将无约束地运行。对晋升锦标赛的改革必须是重新设计，而非废除。

校准能力所要求的重新设计是明确的：在现有的指标达成之外，增加将“准确性”和“问题的早期识别”作为获得积极奖励的行为。这意味着，一名发现辖区内发展中的问题并准确报告的官员——在问题达到无法管理的阈值之前——将获得职业收益而非惩罚。这意味着，一名因拒绝授权会通胀数据的 LGFV 借贷而导致 GDP 增长数据低于同僚的官员，相对于那些同僚不会面临职业劣势。这意味着评估标准应明确区分“在准确信息条件下的表现”与“在管理过的信息条件下的表现”，并独立于绝对产出水平来奖赏前者。

障碍主要不在技术层面——评估标准可以写进这些维度。障碍在政治层面：目前的标准反映了中央真正的优先级，而准确性目前在任何操作意义上都不在其中。一个真正想要准确信息的领导层需要通过评估系统而不是通过号召来发出这一优先信号，且信号必须是可信的——即体现在诚实汇报坏消息的官员的职业轨迹中，而不仅仅体现在赞美求真却惩罚求真者的讲话中。

这是校准架构中最深层的改革，也是对当前系统运行逻辑最具威胁的改革。一个真正因准确性而受奖的官员群体，有时也会递交领导层不想听到的信息。当前的架构通过让这种递交面临职业威胁来防止这一点。改革后的架构将允许这一点，代价是领导层的选择将受到接收到的信息的约束。接受这种约束的意愿，是衡量改革是否真诚的根本考验。

3.6 实验性联邦制的复兴

在当前架构下最立即具备可行性的改革是复兴并保护地方实验传统，这一传统在邓小平时代得以发展，在巩固时代被逐步缩小。

逻辑很直接：地方实验产生了治理方法的差异。差异创造了可观察的结果不同。结果的不同提供了让中央识别优劣的信息。由成功的地方实验生成的信息是目前中国治理系统中最具政治公信力的反馈形式，因为它证明了在“中国条件下”的成功，而非从外国系统引入结论。

地方实验所要求的保护不是实验的“许可”——中央始终是决定哪些实验被允许的最终权威。它要求的是对失败实验的“追溯性惩罚豁免”。目前的激励结构意味着，尝试新方法的地方官员如果未获成功，将面临与失败挂钩的风险；而遵循既定方法的官员即便产出平庸，风险也很小。其结果是产生了系统性的常规行为偏差，而这正是校准所需的“差异性”的敌人。

一个正式的《实验治理协议》——指定特定的省份、城市或行业领域作为受保护的实验空间，在其中尝试新方法而不必承担常规的失败后果，评估基于所生成的经验教训而非既定结果——将恢复使邓小平时代自适应治理成为可能的结构性条件。该协议不会移除中央批准实验和推广成功案例的权力。它将改变目前导致地方官员不愿进行实验的激励计算。

将实验许可地理集中在拥有现有行政能力和愿意根据评估标准而非指标标准运行的政治领导层所在的地区，将允许中央在维持监督的同时，生成产生校准信息所需的差异。沿海省份——广东、浙江、江苏——历史上一直是改革的孵化器，其行政复杂性和经济成熟度使其成为治理实验的理想实验室，而行政约束较多的内陆省份则无法维持此类实验。

3.7 当前架构的局限性

3.3至3.6节描述的改革是关于功能性校准架构所要求的提案。它们不是关于当前中国治理系统将实施什么的预测。必须诚实对待两者之间的关系。

上述每项改革都以特定的方式威胁到了当前架构的运行逻辑。受保护的反馈渠道创造了其诚实评估可能与中央战略方向矛盾的制度参与者。可逆决策结构将错误的制度化承认内置于政策过程中。激励改革创造了因递交不悦真相而受奖的官员群体。实验性联邦制产生了可能产生与中央偏好模型不一致结果的差异。每项改革，按照本系列底层的治理工程框架，都是增加系统校准能力的机制。而每一项，按照当前架构的运行逻辑，都是增加分布式权威和诚实反馈的机制——而这正是权力巩固项目旨在压制的东西。

权力巩固不仅仅是个人权力的集中，它是将治理架构从产生发展奇迹的分布式、反馈丰富、实验多样的系统，系统性地重新配置为更加统一、意识形态一致、中央直接领导的系统。这种重新配置的理由并非不理性：分布式系统产生了大规模腐败、各地区政策不一致，以及出现了足以制约中央权威的经济和政治利益集团。巩固解决了这些问题，但代价是牺牲了分布式系统提供的校准能力（尽管该系统存在效率低下）。

当前架构在现有政治约束下能做的虽然有限，但并非微不足道。在既定战略方向内的战术性航向修正可能的，且经常发生。产业政策可以被调整，而无需承认初始设计错误。经济刺激可以被部署，而无需正式修订增长模型。在非敏感领域——数字治理工具、城市规划方法、服务交付机制——的地方实验可以被允许，而不威胁政治架构。这些都是真实的生产力，不容忽视。但它们与上述结构性校准改革有着本质不同，因为它们是在当前架构内运行，而非改变架构。

因此，本节诚实的结论与本系列对俄罗斯的结论一致：功能性校准能力所要求的改革与当前架构的运行逻辑是不兼容的，且当前架构没有内生的机制来产生实施这些改革的政治意志。与俄罗斯不同的是，这并不是因为改革所需的制度底色已被摧毁。中国保留了非凡的制度能力、技术精湛的官员

群体，以及包含自适应治理先例的治理传统。邓小平时代模式证明了中国治理系统可以拥有功能性校准架构。问题在于，恢复该架构所需的条件是否在当前政权的创造能力之内——现有证据表明并非如此，因为所需的改革将削弱权力集中，而这正是当前领导层定义的其任期内的核心工程。

3.8 仍具可能性的空间

前一节的诚实评估不应陷入宿命论。中国不是俄罗斯。改革所需的制度底色并未像俄罗斯那样被摧毁。包含自适应能力、务实修订和分布式实验的治理传统并未被抹去——它只是被压制了，而被压制的传统在条件改变时是可以恢复的。

改变这一计算的条件很可能是外部而非内部的，因为生成系统内部改革压力的内部反馈架构恰恰被校准赤字所损害。校准迫使性变革最显著的潜在来源是经济：一个长期的经济减速期，使得“绩效合法性”论点——即当前架构因其产出而被证明正确——难以维持。人口约束、债务积累、以及因压制产生早期增长的反馈丰富型经济而导致的创新天花板——这些都是无论治理选择如何都会加剧的聚合压力，并最终将使校准赤字的成本在架构能理解的术语中显现。

变革的第二个潜在来源是精英学习。中国治理系统保留了技术精湛的领导层，历史证明，当证据变得压倒性时，他们有能力更新自己的现实模型。1978年的改革开放不是由外部压力强加的，它是由于内部从毛泽东时代灾难性校准失败中吸取的教训驱动的。当前领导层对巩固项目的承诺不一定是永久的，它反映了对分布式能力与政治控制之间权衡的特定评估，这种评估原则上可以在该权衡的成本变得更加明显时被修订。

在当前约束下仍然可能实现的，是芬兰所谓的“吸引式扩展”版本——通过在特定领域取得的具体治理成功，逐步证明更准确的反馈会产生更好的结果。选定城市的数字治理实验、在社会政策执行中管理过的地方差异、特定工业领域的技术咨询过程——这些虽然不是校准架构所需的结构性改革，但如果条件改变，它们是结构性改革可能萌发的种子。任务是维持这些种子而不是让它们被消除，并将其设计为“拥有更多功能性反馈架构的治理能实现什么”的演示——这些证据在政治计算发生转移时，将作为基础而非需要从零开始重建的东西发挥作用。

3.9 改革如何相互强化

本节描述的改革是相互依赖的。受保护的反馈渠道产生准确信息，但这些信息只有在存在可逆决策结构的情况下才能改善结果，通过这些结构，信息可以触发修订，而不是被过滤进现有的轨道。可逆决策结构只有在激励结构奖赏官员触发修订，而不是因承认需要修订而惩罚他们时，才能发挥作

用。激励改革只有在诚实汇报的官员随后不会通过正式评价标准之外的非正式渠道受到惩罚时，才能产生诚实报告。实验性联邦制只有在实验产生不悦结果而受到追溯性惩罚豁免时，才能生成校准信息。

每项改革都是功能性校准能力的必要非充分条件。它们共同构成了治理系统的最低限度可行架构，使系统能以中国挑战所要求的速度更新其现实模型。在当前架构内，每一项都可以以有限的、战术性的方式实施。但在结构上同时实施这些改革，将构成对中国国家运行逻辑的重大重新配置——这种配置以牺牲巩固项目所优先考虑的统一、中央直接领导的权威为代价，换取校准能力的提升。

这种紧张关系并不是解散改革的理由。它是理解这些改革最重要的一点。一个优先考虑“执行”而非“校准”的治理架构，在引导执行的模型准确的领域会产生非凡的结果，而在模型不准确的领域则会产生灾难性的结果。中国的发展奇迹是前者的证据；动态清零的崩溃、LGFV债务的积累、人口政策迟缓且不足的修订，则是后者的证据。只要产生这种架构的基础保持不改，“运动一过头一突发修正”循环将继续产生其特征性的跟踪。改变的只是下一次修正的领域、规模和代价。

下一节将描述阻碍改革的“免疫系统”——不是将其作为一个抽象的政治障碍，而是作为其运行逻辑使校准改革产生结构性威胁的特定制度和文化架构。理解免疫系统是诚实评估上述改革在何种条件下可能成为可能的先决条件。

4. 政治免疫系统：控制保存本能 (The Control Preservation Imperative)

4.1 中国免疫系统的独特之处

本系列研究过的每一个治理系统都形成了针对权力再分配的特征性防御机制。芬兰的免疫系统是“稳定偏好”——一种嵌入在使国家易于治理的制度和文化中的被动结构性倾向。瑞典的免疫系统是“满足的胜任力”——整体表现足够强劲，以至于过滤掉了扰动的信号。法国的免疫系统是“雅各宾式奇观”，它在改革能量产生制度变革之前将其转化为意识形态冲突。巴西的免疫系统是“大中间派”(Centrão)——一个吸收任何改革冲动并将其转化为交易性租金的热力学沉降池。美国则拥有“否决工业复合体”——一个从僵局中获利的专业生态系统。

这些免疫系统都非常强大。它们主要通过减缓、吸收或转移改革来发挥作用，而不是主动摧毁改革所需的条件。芬兰的潜在改革者面临的是制度惯性和文化阻力。法国的潜在改革者面临的是针对其提议而动员起来的奇观。这些都是严重的障碍，但改革者可以通过耐心建立联盟、精细的设计，以及本系列在每个背景下识别出的“木马机制”来应对。

中国的免疫系统本质上是不同的。它不是被动的惯性，也不是一种政治沉降池。它是一套主动的、受到严密资源支持的、嵌入在架构中的**控制保存本能 (Control Preservation Imperative)**。它不仅抵抗改革，更将产生功能性校准所需的每一个先决条件——分布式的权威、诚实的反馈渠道、受保护的异议空间、以及对中央战略方向的实验性挑战——识别为对系统生存逻辑的结构性威胁。

4.2 控制与稳定的混淆

中国免疫系统的核心逻辑在于将“控制”与“稳定”完全等同。在当前架构的学说中，只有通过中央对政治、信息、经济和人口变量的直接、统一控制，才能维持稳定。任何分布式的权威或未经管理的反馈都被视为动荡的前兆。

这种等同在历史上并非没有根据。中国历史上每一次中央权威的削弱都伴随着动乱、割据和外辱。1989年的动荡被解释为未能充分控制信息的后果，而非未能解决根本性的校准赤字。20世纪90年代的腐败被解释为各部门和地方权威过于分散的后果。从这个视角看，权力巩固项目不仅是为了权力本身，也是作为一种治理改良——通过消除产生功能性障碍的竞争性权力中心来“修复”系统。

这种免疫反应的悲剧在于它是自我消解的。为了维持短期稳定，系统压制了产生长期韧性所需的校准机制。这种免疫系统越是成功地压制了对政策失败的早期反应，校准赤字积累的压力就越大，直到最终迫使修正发生的事件——例如房地产危机或大流行病的失控——反而在更大规模上威胁到系统想要保护的稳定性。

4.3 为什么校准改革被识别为威胁

第三章描述的功能性校准改革被当前系统的免疫系统准确地识别为生存威胁。理解这一点对于任何关于中国改革可能性的诚实评估都至关重要。

受保护的反馈渠道：建立一个旨在提供诚实、有时是批判性评估的独立机构，意味着创造了一个其合法性不源于“步调一致”的制度行为者。在目前的控制架构中，这种机构会被视为潜在的挑战者，是可能被政治对手利用来破坏中央战略方向的工具。

可逆的决策结构：将“修订”内置于政策程序中需要公开承认模型的局限性。在目前的合法性框架中，这种承认会被视为软弱，会鼓励进一步的挑战。免疫系统宁愿选择一个延迟但代价高昂的逆转，也不愿选择一个及时但会显示模型并非全知全能的修正。

激励机制改革：奖赏官员递交坏消息意味着承认地方层级的现实比中央模型更权威。这会削弱中央用于指导庞大官僚机构的行为约束，而免疫系统认为这种约束是防止系统陷入地方主义和腐败的唯一手段。

实验性联邦制：允许地方实验产生差异意味着失去了中央对治理叙事的控制。如果一个省份成功了而其他省份失败了，这种对比产生的政治压力是目前控制架构无法容忍的。

这些并不是改革者面临的意外障碍。它们是免疫系统的设计特征。它们的存在是为了防止系统偏离预设的轨道。其结果是产生了一个在执行轨道内表现非凡，但在改变轨道方面具有结构性无能的系统。

4.4 被管理的认识论

中国免疫系统最精密的一面是其对认识论环境的管理。正如第二章所讨论的，这已经超出了传统的审查，演变为对国家产生和处理自身治理数据能力的系统性降解。

这种免疫反应之所以强大，是因为它表现为一种技术上的精进。利用人工智能、大数据和预测性监控来“管理现实”被视为治理的现代化，而不是功能的损失。然而，这种管理创造了一个反馈闭环：系统在管理过的数据上训练出的模型，生成了进一步管理现实的指令，从而掩盖了现实领土。

对免疫系统而言，这种闭环是理想的。它创造了秩序的表象，并消除了直接挑战。但对治理架构而言，它是毁灭性的。它确保了校准赤字不仅在扩大，而且变得越来越不可见，甚至对那些负责管理它的人也是如此。这是一种“认识论上的自我中毒”，免疫系统为了保护身体免受异议的侵扰，而逐渐切断了身体感知痛苦的能力。

4.5 与其他系统的对比：阈值效应

芬兰或法国等国家的免疫系统运作会产生“摩擦”。它们让变革变得缓慢、令人沮丧且充满了政治妥协。但由于它们没有摧毁反馈回路，它们允许系统通过一系列的小调整来应对不断扩大的赤字。

中国的免疫系统运作会产生“阈值效应”。由于它压制了增量式的反馈和调整，它确保了校准赤字只能通过剧烈的爆发来清算。控制被维持直到无法维持的那一刻，然后这种失控是与之前压制的深度和时长成正比的。这就是“运动一过头一突发修正”循环的潜在逻辑。

这就是为什么中国的免疫系统在所有研究案例中是最危险的。其他系统通过降低效能来换取稳定。中国的系统通过增加系统性崩溃的风险来换取近期的完美控制。

4.6 结论：一个无法自纠的模型

控制保存本能不是一种政策选择，它是一个已变得过于成功的治理架构的结构产物。它成功地排除了所有可能触发自纠的内部机制，因为它将所有此类机制都准确地识别为对其权威的稀释。

这就留下了一个最终的悖论：中国政府拥有非凡的能力来解决它所定义的问题，但由于校准赤字和免疫系统的共同作用，它越来越无法可靠地更新其关于“问题是什么”的模型。

该系统不是不稳定的。领土是不稳定的。防止对两者进行准确评估的认识论反馈崩溃是不稳定的。每一种都是延迟的不稳定来源，其成本随着延迟的持续而复合成长。

保护当前架构免受校准改革影响的免疫系统，从这个意义上说，是校准赤字自我复制的主要机制。它压制了那些本能减少赤字的改革，从而确保赤字继续产生阈值事件，最终在危机而非深思熟虑的条件下迫使改革发生。控制保存本能保存了控制，直到它失效的那一刻；那一刻控制的丧失，与此前校准失败的持续时间和深度成正比。

这是中国报告对本系列的最终诊断贡献。本系列中其他国家的免疫系统——芬兰的稳定偏好、巴西的大中间派、法国的雅各宾奇观——虽然抵抗改革，但并不摧毁改革的条件。中国的控制保存本能则定性不同：它本身就是其所捍卫的校准赤字的主要驱动力，确保了国家的现实模型随着免疫系统成功阻止改革的每一年，都进一步偏离现实。下一节将描述一个除了修正自身模型之外，能做成任何事情的系统的结语。

5. 具体的第一步：实验性治理协议 (The Experimental Governance Protocol)

5.1 可能性的逻辑

没有“第一步”的框架只是思想实验。本系列之前的每份报告都提出了具体的、可实施的机制，旨在相关系统的政治约束内解决所诊断出的赤字。芬兰的“未来影响评估”、法国的“自适应整合领土”、巴西的“算法绕道”——每一项设计在现有的政治架构内都是可行的：初始规模适度，在要求更广泛的承诺之前能够证明价值，并且与实施者（即相关参与者）的激励结构保持一致。

中国报告面临着不同的挑战。前几节已经坦诚地指出：功能性校准能力所需的改革与当前架构的运行逻辑在结构上是不兼容的。第四节描述的免疫系统不仅是减缓校准改革，它准确地将其识别为威胁并做出相应反应。因此，在这一背景下可行的“第一步”提案与芬兰或法国的提案在本质上有所不同：它们不是最终将实现完整改革方案的“转型架构”的前沿；它们是在当前约束下所能达到的最大极限——是如果条件改变，更深层改革可能从中萌发的种子，而不是能保证产出这种改革的机制。

5.2 实验性治理协议 (EGP)

提案是建立**“实验性治理协议” (Experimental Governance Protocol, EGP)**。这并非通过法律变动来实现，而是通过行政指令：由国务院正式指定一系列省份、城市或特定的产业领域作为“受保护的实验空间”。

在这些空间内，地方官员被明确授权实施与中央当前战略方向不同的治理方法，前提是这些方法要遵循一套严格的、预先定义的实验协议。

该协议的三个核心支柱是：

- 追溯性惩罚豁免：**参与 EGP 的官员将获得正式保证：如果实验未能达到预期结果，只要该实验是按照商定的协议执行的，其职业生涯就不会受到负面影响。其评估将基于所生成的经验教训的质量（即信息的准确性），而非结果是否与预设指标一致。
- 独立验证层：**实验结果的评估不应由实施实验的行政官僚机构负责，也不应由其上级负责。评估应由一个独立的验证小组进行，该小组由学术专家、技术型官员以及来自非参与地区的代表组成，其唯一任务是准确记录所发生的情况。

3. **受控的知识汇聚：**实验产生的发现不应直接进入公共领域，也不应通过官僚报告系统向上呈报。相反，这些发现应进入一个“受保护的知识库”，仅供最高决策层及参与实验的实体访问。这降低了实验结果被政治化或被视为对中央权威公开挑战的风险。

5.3 为什么这具备可行性

EGP 旨在绕过（而非消除）第四节描述的免疫系统。它利用了系统内仍然存在的、被压制但尚未被摧毁的组件：

- **利用“试点”传统：**中国的领导层对“试点”有着深刻的体制记忆。EGP 将此包装为这一传统的现代化回归，而不是西方式的改革。它向免疫系统发出信号：这不仅不是威胁，反而是通过在将风险推向全国之前先进行局部测试来“降低”风险。
- **低政治能见度：**通过限制实验结果的传播（受控的知识汇聚），EGP 避免了那种会导致中央因担心失控而关闭实验的公开对比。它允许中央在不公开承认模型可能存在缺陷的情况下，获得校准模型所需的信息。
- **与绩效合法性对齐：**EGP 被呈现为一个提高行政效率和解决具体治理难题（如人口老龄化或地方债）的工具。它并不声称要重新分配权力；它声称要生成让权力运行更有效的信息。

5.4 该机制的局限性

我们必须明确 EGP 不能做什么。它无法解决权力巩固项目产生的根本性校准赤字。只要系统的最高层仍然根据受损的反馈运行，局部的、受保护的实验产生的信息就只能产生局部的、受访的效益。

EGP 真正的价值在于**保存**。它是为了在当前架构最不愿容忍差异的时刻，在系统内部保留住“差异”和“诚实反馈”的火种。它是为了维持一小群了解如何运行实验、如何处理坏消息、以及如何根据证据而非指标进行管理的官员。它是为了证明——即使规模很小——一种不同的反馈架构确实能产生更有效的发展成果。

5.5 第一步之后的路径

EGP 产生的证据——如果它被允许运行——将成为未来更深层改革的基石。当校准赤字的成本变得如此高昂，以至于免疫系统的防御开始瓦解时（这通常发生在严重的经济或社会压力下），系统将需要现成的、经过验证的替代模型。EGP 提供的就是这些模型。

它将“表现优异的治理”与“高度集权的控制”脱钩。它提供了一个现成的证明：恢复实验性、分布式和反馈丰富的治理模式（这些正是成就发展奇迹的特质），不仅不会导致动荡，反而是应对21世纪复杂挑战的唯一途径。

5.6 结语：向事实求是

本系列关于中国的报告始于对“校准赤字”的诊断——这是国家的现实模型与它所宣称代表的现实领土之间的脱节。它追踪了这一赤字如何通过晋升锦标赛、地方债、认识论崩溃和人口约束而产生。它描述了功能性校准的要求，以及会阻碍这些要求的免疫系统。

“实验性治理协议”是一个适度的第一步。它并不试图一次性修复整个架构。它试图在架构内为“事实”创造一个避风港，使其免受“步调一致”这一首要任务的挤压。

朝着这一目标迈出的第一步，指向了一个能够恢复曾经创造发展奇迹的校准能力的中国治理系统。这种系统能够以发展奇迹所带来的复杂性所需的准确度，来感知其经济、人口、环境和地缘政治地位；能够在其行动的成本超过变革的成本之前修订战略；能够将呈报不悦真相的人视为资产而非威胁。

那样的治理系统并非目前的系统。它与目前的系统相去甚远。但它是具有辨识度的“中国式”系统——它扎根于“实事求是”的务实传统，植根于建立经济特区作为校准工具的先例，以及“摸着石头过河”的经验。当前系统与那个愿景之间的距离，就是权力巩固项目所付出的代价。第一步就是开始缩小这一距离，在目前唯一可用的领域，利用目前约束所允许的机制——并以一种能够积累证据的方式进行，使得这些证据的价值在显现时，不会再次被产生赤字的旧架构所吞噬。

6. 结语：模拟与领土 (The Simulation and the Territory)

6.1 中国所成就的

对中国治理成就的诚实描述，是诚实审视其校准赤字的必要起点，因为两者密不可分。在四十年的时间里，中国实现了一个治理系统在人类历史上从未以如此规模和速度实现的成就：将14亿人口的社会从贫困转变为接近发达世界的生活水平。支撑这一转型的基础设施——跨越大陆的高铁网络、吞吐量超过世界其他地区总和的港口设施、以及在三十年内扩建规模超过任何其他文明史上总和的发电能力——是协调的国家能力在与连贯的发展模型对齐并数十年如一日地坚定执行时，所能创造出的真实丰碑。

这并非边缘性的观察。这是任何治理框架在提供诊断之前必须考虑的最重要的一点。一个无法承认威权协调在特定条件下能达到的效果的框架，不是一个严肃的框架。发展奇迹是真实的。脱贫是真实的。数亿人日常生活的物质改善是真实的，而这些都是本系列所描绘的“治理选择”的产物。

6.2 模拟系统

这种同样的协调能力也被用于构建世界上最先进的治理“模拟系统” (Simulation)。中国构建了国家曾建造过的最精密的社会模型之一。数据系统、监控基础设施、行政报告机制、计划机制——它们共同构成了一个广度极广、技术极其先进的关于现实的表征。在许多领域，这个模拟系统表现得异常出色：它能精确地引导物流，能以非凡的效率管理数字支付，能在危机中以前所未有的广度动员人口。

然而，正如本报告所记录的，“校准赤字”就是这个模型与它所代表的真实“领土” (Territory) 之间发生的进行性脱节。

- 晋升锦标赛过滤了进入模拟系统的信息。
- 地方政府融资平台 (LGFV) 架构创造了模型财政表征中未捕捉到的真实负债。
- 认识论反馈崩溃缩减了模型构建者可用的视角范围。
- ** “明哲保身” **确保了那些私人知识与模型表征偏差最大的人，最不可能揭示这种偏差。

模拟系统以非凡的复杂性和力量运行。而现实领土无论模拟系统如何描述，依然按照其自身的规律存在。

6.3 最终的清算

“运动一过头一突发修正”循环是领土对模拟系统的反击。当模拟系统坚称动态清零是成功的，而领土却显示病毒已失控且社会契约已断裂时，模拟系统便会崩溃。当模拟系统显示房地产驱动的增长是可持续的，而领土却显示债务已无法偿还且建筑工地的吊车已停工时，模拟系统也必须做出调整。

校准赤字并不意味着系统是无效的。它意味着系统越来越依赖于那些只有在不再能通过行政手段管理时才会显现的信号。模拟系统越强大，它就越能长期地压制来自领土的信号——而最终的清算也就越剧烈。

6.4 备选架构

本系列提供的框架不是为了预测模拟系统何时会与领土发生最终碰撞。它的目的是提供一种“备选架构”，使这种碰撞能够被管理，而非仅仅被承受。

如果条件改变——无论是通过由于绩效合法性侵蚀带来的精英学习，还是由于足够深重的校准失败所引发的内部危机——本报告所设计的方案是现成可用的，而非在压力之下仓促开发出来的。本报告描述的受保护反馈渠道、可逆决策结构、激励机制改革以及实验性治理协议，并不是立即实施的处方。它们是一个可能未来的架构——一个中国治理系统可以拥有的未来，因为邓小平时代的模式证明了中国系统具备这种架构的能力，且这种能力与非凡的国家成就并存。

模拟系统最终必须面对领土。两者之间的差距衡量了校准赤字所付出的代价，也衡量了修正这一赤字所需的努力。本系列框架提供的不是关于这场碰撞何时到来或以何种形式到来的预测。相反，它提供了一张架构地图，允许这场碰撞被引导而非仅仅是被忍受——从而使从修正中浮现的治理系统，能够从失败中学习，而不仅仅是在失败中幸存。

发展奇迹是人类治理的重大成就之一。本系列无法回答——但领土最终会以它自己的方式回答——的问题是：创造了这一奇迹的治理系统，是否能创造出中国及世界所需的下一个奇迹。

附录 A：价值体系与政策思维 —— 中国背景指南

关于本附录的说明

本报告正文避免使用发展心理学或文化理论的专业术语，而是采用治理架构、校准能力和反馈设计的语言。本附录为希望了解中国治理中深层价值体系动态的读者提供了一个补充视角。它是可选内容，但它使报告的底层逻辑完全透明化，并直接回答了本系列资深读者会提出的问题：一个像中国

这样在制度上如此完善的治理系统，是如何走向校准边界条件的？

A.1 基本洞察

不同的机构和政治文化在思考治理、资源和变革时，往往有不同的重心。这些不是人格类型或党派所属，尽管它们与两者都有微弱的相关性。它们是底层的**价值体系**——构建什么是真实的、合法的和重要的方式。每种价值体系都代表了对特定生活条件的连贯反应。没有哪种体系在绝对意义上是“更好”的。每种体系都有在特定条件下显现的特征性优势，以及在其他条件下显现的特征性盲点。

此处使用的框架借鉴了**螺旋动力学 (Spiral Dynamics)** 整合理论。以下是与中国治理最相关的体系简图：

- **蓝色 (秩序/纪律)**：核心价值是稳定性、合规性、权威和等级制度。它通过建立明确的规则、正当的过程和对集体标准的忠诚来应对混乱。
- **红色 (力量/意志)**：核心价值是果断、主导、行动和生存。它通过个人英雄主义、意志力和突破障碍的原始能力来应对威胁。
- **橙色 (成就/战略)**：核心价值是效率、竞争、物质进步和功能性的成功。它通过科学思维、战略规划 and 基于结果的逻辑来应对停滞。
- **黄色 (自适应/系统)**：核心价值是系统思维、自适应能力、灵活性和功能的集成。它通过拥抱复杂性、分布式反馈和根据证据不断更新模型来应对复杂系统。

A.2 中国治理的价值组合

中国奇迹的非凡之处在于它能够将**蓝色、红色和橙色**整合成一个高度协调的整体：

- **红色**提供了突破制度阻力、快速动员资源和设定极具野心目标的原始冲动（如早期的基础设施激增）。
- **蓝色**提供了党和国家的组织结构，确保了这种冲动不会演变为军阀割据，而是通过统一的行政层级进行。
- **橙色**提供了驱动经济发展的务实主义，利用市场机制、衡量增长指标并追求技术优势。

这种“蓝—红—橙”的组合产生了世界历史上最强大的**执行架构**。它能建高铁、消灭绝对贫困并协调全球供应链。

A.3 校准赤字的来源

当这种架构在没有足够**黄色 (自适应)** 机制制衡的情况下运行时，校准赤字就会出现。

在“蓝—红—橙”系统中：

- **蓝色**将异议视为对秩序的威胁。
- **红色**将修正视为软弱的表现。
- **橙色**通过确定的指标来衡量成功，这创造了“晋升锦标赛”中系统性的激励：即管理报告而非管理现实。

黄色的缺失意味着虽然系统在“做事情”上非常出色，但它缺乏结构性的能力来怀疑“正在做的事情是否仍然是正确的”。它拥有强大的引擎（红/橙）和坚固的底盘（蓝），但缺乏对领土变化的灵敏感知系统（黄）。

A.4 权力巩固的影响

邓小平时代的系统代表了中国最接近整合**黄色**自适应能力的时期。虽然中央仍保持着“蓝—红—橙”的控制，但它有意识地为地方层级的“黄色”实验创造了空间。经济特区、农村改革和“摸着石头过河”的方法论本质上都是黄色机制——它们拥抱差异，利用分布式的反馈进行校准，并允许模型根据现实证据进行修订。

当代的**权力巩固项目**，在价值系统术语中，是一个系统性地强化“蓝色”和“红色”控制并牺牲“黄色”自适应能力的过程。为了消除分布式权威带来的副作用（如腐败、不一致），系统也一并消除了校准所需的反馈渠道。

- **红色/蓝色**的复兴维持了行动的果断（如动态清零的早期成功），但由于缺乏**黄色**的校准，导致系统在环境改变后无法及时转向。
- **橙色**的成就导向仍以定义的指标衡量绩效，但这在当前架构下反而加强了伪造数据和粉饰报告的动力。

A.5 功能性校准的要求：向黄色集成

报告第三节描述的改革（受保护的反馈渠道、可逆的决策结构、实验性协议）本质上是要求在当前的架构中重新注入**黄色**的自适应能力。

这并不意味着要用黄色取代蓝、红或橙。那将导致系统瓦解。真正的挑战是**集成**：

- 保留**蓝色**的制度连贯性。

- 保留**红色**的果断执行力。
- 保留**橙色**的战略驱动力。
- **但**，必须整合**黄色**的自适应机制，以确保前三者的力量是被一个不断根据现实领土校准的模型所引导的。

当前的障碍不是中国缺乏黄色思维——中国的技术专家、科学家和许多地方官员都具备这种素质。障碍是结构性的压制：当前的“蓝—红—橙”架构认为黄色所要求的分布式反馈是对其控制权的威胁。

结论

恢复校准能力，在价值系统术语中，意味着恢复曾成就发展奇迹的集成状态：即利用橙色的务实提供动力，蓝色提供架构，并重新允许适度的黄色自适应实验提供校准反馈。如果无法实现这一集成，系统将继续在“蓝—红—橙”的惯性下加速，直到其现实模型与领土的差距再次引发无法被模拟系统掩盖的危机。

附录 B：国际类比与先例

本报告中的建议并非没有先例。以下示例展示了曾应对过与中国类似校准挑战的治理系统——这些示例并非作为直接采纳的处方，而是作为证据，证明自适应治理的功能要求可以在非西方自由民主传统的治理传统中得到满足。

B.1 新加坡：拥有更功能化反馈的受控发展

新加坡是与中国最具参考价值的比较对象，恰恰因为它不是西方民主国家，不能被斥为异质模型。新加坡的治理系统与中国一样，致力于强有力的中央领导、务实而非意识形态化的经济政策，以及一党主导的政治架构。

将新加坡的治理与中国当前轨迹区分开来的是，新加坡在政治架构内刻意保留了技术官僚反馈机制：

- ****经济发展局（EDB）****有能力对产业战略提供诚实的评估。
- **公务员系统**拥有基于展示的能力而非政治立场的功绩晋升传统。
- **明确的规范**：政策失败应被分析和纠正，而不是归咎于外部原因或进行压制。

新加坡在自由主义意义上并不是一个自由社会。它的反馈架构在某些方面受到限制，但它证明了一个威权协作系统可以建立并维持诚实的内部反馈回路，使其模型与领土保持同步。

B.2 20世纪70年代的日本：作为校准机制的产业协作

20世纪70年代的日本——即其处于发展巅峰且常被西方视为威胁的时期——提供了一个关于“非竞争性反馈”的案例。通商产业省（MITI）并不是通过直接指令来运行经济，而是通过一种被称为**“行政指导”**的过程，这是一种持续的、机构化的反馈循环，连接了国家计划者、工业领袖和技术专家。

这种架构允许日本在面对1973年石油危机等外部冲击时，比大多数西方民主国家更快速、更彻底地重新校准其整个工业基础。这种校准能力并非源于选举竞争，而是源于国家与私营部门之间存在的受保护的信息共享渠道，这些渠道确保了决策依据的是实时的工业现实，而非中央的愿望。

B.3 18世纪的普鲁士：作为反馈渠道的司法独立

虽然在时间上较为久远，但普鲁士在腓特烈大帝时期的经验提供了一个极具针对性的例子：即一个高度集权的威权国家如何通过建立独立的司法机构来增强其校准能力。

腓特烈意识到，如果他的官僚机构可以随意违抗法律，他就无法获得关于政策效果的真实信息。通过建立一个可以裁决涉及国家案件且不受政治干预的法院系统，他创造了一个**法律校准渠道**。当公民在法院起诉国家并获胜时，这不仅仅是法律正义的体现，更是一个功能性的信号，告知中央政府其行政机器在何处产生了摩擦、不公或低效。这是一种在不稀释中央主权的情况下，利用法律体系作为反馈传感器的行为。

B.4 越南：小规模实验主义

越南目前的治理路径提供了与中国“巩固项目”的当代对比。越南与其北方邻国共享许多制度基因，但它保留了（部分是由于其实际的权力分散）更高水平的地方实验主义和对政策失败的公开承认。

越南较小的规模使得这种比较并不完美，但其更愿意承认并比中国更迅速地纠正农业政策失败，以及在经济整合上采取更具实验性的方法，这些都表明：当前中国模型的特定校准失败并非治理模型本身的固有缺陷，而是源于那个逐步收窄了反馈架构的特定“巩固项目”。

B.5 邓小平时代：最重要的中国先例

最重要的先例是报告 3.2 节中作为改革论点核心的部分：中国治理系统自身的历史。

- **经济特区**并非外国进口。
- **乡镇企业**并非西方建议。
- ** “摸着石头过河” **的方法论并非借用的概念。

这些是中国官员针对中国治理挑战而开发的本土创新。它们之所以能产生非凡的结果，是因为它们在制度形式上体现了当前系统已逐步压制的校准原则。

对校准改革最强有力的论点不是它会让中国变得像其他国家。而是它会让中国变得更像那个创造了发展奇迹的中国——理解这一点，是诚实评估巩固项目中失去了什么、以及需要恢复什么的先决条件。

附录 C：治理与工程的联系

C.1 建筑学基础

本报告借鉴了更深层次的研究成果：《治理即工程》(Governance as Engineering) 系列。该系列是一套正式的分析报告，利用控制理论、信息论和控制论的标准数学方法，将治理机构建模为反馈控制系统。该系列具有很强的技术性；本附录用非技术性语言总结了其核心发现，并展示了这些发现如何支撑了“校准赤字”的诊断。

C.2 五篇论文与校准赤字

论文 I —— 《治理稳定性模拟器》证明了运行在聚合信号上的集权治理系统会破坏“空间信息”。这是本报告论点（即晋升锦标赛对上报结果——GDP 增长、稳定指标、合规指标——的系统性聚合破坏了信息中的空间和时间变异性，而这些变异性本可以允许中央察觉到上报绩效与实际情况之间的脱节）的正式基础。地方政府融资平台（LGFV）债务架构正是空间信息被破坏后的精确表达：中央的聚合财政图景看起来很稳健，而分散的现实却在积累被聚合过程所掩盖的负债。

论文 II —— 《作为稳定性的分形性》证明了在面对同时发生的快、中、慢速扰动时，没有任何单一尺度的控制器能够稳定系统。中国治理系统面临着极快（社交媒体驱动的动荡）、中速（经济周期、房地产转型）和慢速（人口衰退、债务积累）的扰动。权力巩固项目通过削弱地方和中间层级的自适应能力，将系统推向了单一尺度的控制器。模拟结果显示，这种架构在短期内表现出极高的稳定性，但代价是极易受到跨尺度共振的影响，即慢速的压力（如人口结构）会突然放大中速的失败（如养老金偿付能力），并触发极快的危机（如社会抗议）。

论文 III —— 《主权作为一种通信协议》探讨了将“一致性”作为主要治理指标的信息成本。当一个系统奖赏“步调一致”超过奖赏“准确性”时，它实际上是在强制执行一种会导致信息熵损失的通信协议。在这一框架下，校准赤字被测量为“协议开销”：即为了维持政治一致性的表象，而不得不从行政通道中过滤掉的真实世界信息的总量。

论文 IV —— 《校准的拓扑结构》提供了证据，证明如果反馈回路经过的层级超过了其试图控制的过程的复杂度，那么反馈就会变得具有破坏性。在中国的背景下，这解释了为什么将所有反馈（如地方土地纠纷或环境抗议）都引向北京的尝试，反而会导致校准恶化。中央变得信息过载，而能够解决这些问题的局部自适应机制却因缺乏自主权而萎缩。

论文 V —— 《算法治理边界》建模了数字监控对决策质量的影响。其核心发现是“观察者效应”的宏观版本：当一个系统知道自己正受到完美的监控时，它就会改变自己的行为，使其符合监控者的模型，从而使该模型无法再捕捉到系统的真实状态。这直接支撑了关于“算法现实扭曲场”的论点：国家在数字管理上的成功，反而降解了其治理所需的真实数据。

C.3 总结：带宽与可观测性

从工程角度看，“校准赤字”可以被定义为**低可观测性与高增益控制**的结合。

- **高增益 (High Gain)**：指系统对检测到的信号做出强大、迅速、大规模反应的能力（执行力）。
- **可观测性 (Observability)**：指从系统外部输出（数据、报告、观察）推断系统内部状态（现实领土）的能力。

中国的治理系统拥有极高的增益——它能动员数亿人并建设横跨大陆的基础设施。但由于前述机制，它的可观测性却在下降。中央的“带宽”——即处理中国这种复杂度的系统所产生的多重、同步、多时间尺度信号的能力——正受到行政过滤和政治一致性要求的限制。

一个拥有高增益但低可观测性的控制系统是危险的。它对自己能探测到的信号反应剧烈——而正如晋升锦标赛分析所示，这些信号被系统性地扭曲为好消息——而它无法察觉到模型与现实的偏离，直到这种偏离产生的信号大到足以穿过滤架构。届时，对检测信号的高增益反应——即“突发修正”——是与累积的偏离成正比的，而非与最优修正成正比，因为系统一直无法增量地检测并解决这些偏离。

这就是“运动—过头—突发修正”循环在工程层面的表述。第 5 节提出的“实验性治理协议”在控制理论术语中是一种**可观测性增强**：即通过创造能够产生对比信息的不同方法，增加治理系统对其领土感知的维度。而“绩效准确性审计”则是一种**增益校准**：即通过减少系统自述中的系统性偏差，防止高增益反应被引导至一个扭曲的现实模型而非现实本身。

附录 D：预期的异议

D.1 “西方框架不适用于中国”

这是专业读者最可能提出的异议，值得进行最审慎的回应。此异议有两个版本。第一个是**经验性**的：中国的治理历史、文化传统和制度逻辑与开发该治理框架的西方系统截然不同，因此框架中的概念（如辅助性原则、分布式权威、反馈架构）无法准确翻译到中国语境中。第二个是**规范性的**：该框架的处方反映了西方自由民主价值观，而中国的治理传统并不认同这些价值观，也没有义务采纳它们。

针对经验性版本的回答，体现在本报告始终致力于将其处方置于中国治理先例而非西方模式中。邓小平时代的实验性联邦制、“摸着石头过河”的方法论、经济特区模式——这些都不是应用到中国的西方概念。它们是中国创新，本报告识别了它们的功能逻辑，并主张恢复这些逻辑。本报告描述为必要的反馈架构并不是西方的；它是中国治理系统自身开发、并已证明能产生非凡结果的反馈架构。

针对规范性版本的回答，体现在本报告始终将其处方置于**功能必要性**而非道德偏好的框架内。本报告并未主张为了民主本身的价值而推行民主。它主张的是：中国自身设定的目标——如维持长期增长、应对人口萎缩、管理复杂的债务环境——在功能上需要比目前更高水平的校准能力。这种校准能力需要诚实的反馈和战略修订的能力。如果目前的架构不能提供这些，那么它将无法实现其自身定义的目标，而这与任何外部规范性评判无关。

D.2 “本报告低估了中央控制的必要性”

这种异议认为，报告对分布式权威和自适应反馈的强调，忽视了中国为何在过去十年中转向权力巩固的现实原因。这一观点主张，邓小平时代的分布式模式导致了灾难性的腐败、各省份间的政策割据，以及出现了足以挑战国家稳定和统一的替代性权力中心（如大型科技公司或地方利益集团）。从这个角度看，权力的巩固是对分布式治理产生的功能失调的一种必要且理性的纠正。

本报告承认这一论点的力量。第四节关于“控制保存本能”的讨论明确指出，巩固项目并非非理性的。它是对真实治理挑战的真实回应。然而，本报告的诊断指出：在解决分布式治理产生的副作用（腐败、不一致）的同时，巩固项目也一并消除了分布式治理带来的核心效益——即**校准能力**。

目前的架构并不是在“控制”与“腐败”之间做选择，而是在“完美的行政控制”与“对现实的感知能力”之间进行权衡。目前的轨迹是向前者倾斜，代价是后者。本报告主张的不是回到 20 世纪 90 年代那种无节制的分布式状态，而是转向一种**受保护的、功能性的分布式状态**，这种状态能提供校准所需的反馈，同时又不产生巩固项目旨在消除的那种掠夺性副作用。

D.3 “中国通过大数据和 AI 已经解决了反馈问题”

这种技术决定论式的异议认为，中国不再需要传统的反馈架构（如异议、试点或实验），因为它拥有数字监控、预测性算法和实时行政数据。这一观点主张，国家现在的感知能力是直接的，不需要经过中介，因此传统的“校准赤字”已被技术手段所弥补。

第二节和附录 C 对此进行了技术性的反驳。数字监控并不是中性的观察者。当一个系统知道自己正受到监控时，它就会改变自己的行为以符合监控者的模型（观察者效应）。由于行政数据是由处于“晋升锦标赛”激励结构下的官员生成的，数字工具只是以更快的速度捕捉到了经过扭曲的数据。

AI 并不能通过计算消除校准赤字，反而会将其自动化。如果在被过滤的数据（反映了中央“想看到的”而非“现实存在的”）上训练模型，那么这些模型只会产生一个数字化的、更加完美的“模拟系统”，而这个模拟系统与“现实领土”的脱节只会越来越严重。技术增强了**执行**，但如果不改变底层的激励结构，它会进一步削弱**校准**。

D.4 “这些改革在政治上是不可能的”

这种异议认为，本报告描述的处方（如受保护的反馈渠道或激励机制改革）在目前的政治环境下是不可想象的，因为它们直接违背了当前领导层的核心议程。

本报告对此表示同意。第三节和第五节明确区分了“功能性要求”与“政治预测”。本报告并不是在预测这些改革将会发生；它是在指出，为了使系统继续实现其目标，这些改革是**必须**发生的。

本报告的作用不是提供政治占卜，而是提供**治理工程的蓝图**。如果目前的架构被证明无法应对未来的挑战（如人口或债务危机），那么系统将需要一套现成的、经过验证的替代逻辑。第五节提出的“实验性治理协议”正是为了在这种受限的环境下提供一个起点。它们被设计为在中国现有的行政架构内运行，而不是要求对其进行全面的重新设计。

D.5 “校准改革会使系统不稳定”

这一异议是最严肃的，也是本报告最慎重对待的。第四节对“控制保存本能”的分析承认，将“稳定”与“控制”混为一谈——即认为分布式的权威是不稳定的前兆——在自身逻辑内并非不合理。改革开放时期真实的治理失败（包括大规模腐败和足以制约中央权威的利益集团的出现）为巩固项目的担忧提供了真实的背景。

本报告的回应是结构性的，而非否定性的：一个通过压制反馈架构来维持短期稳定的治理系统，并没有真正实现稳定。它只是在**推迟不稳定**——用早期航向修正带来的可控扰动，换取了后期“阈值事件”带来的可能无法控制的扰动。LGFV 债务的积累是不稳定的。人口轨迹是不稳定的。认识论

反馈的崩溃是不稳定的。每一个都是延迟不稳定的来源，其成本随延迟的持续而复合成长。校准改革不是对稳定的威胁，它们是稳定性能延续到当前政治周期之后的先决条件。

附录 E：关于作者与方法论

作者说明

本报告是基于对横跨多个大洲的治理系统的比较研究而撰写的，但作者并非来自中国体制核心。作者不是中国人，不住在中国，也不自称拥有在中国治理体系内部生活的权威经验。这里提供的视角源于对复杂性科学、发展心理学（螺旋动力学）、治理理论以及机构控制论设计的长期投入。作者深信，最有价值的诊断有时来自于被诊断系统之外，因为外部视角可以提出内部人士已经习以为常或不被允许提出的问题。

与制度权力的距离既是一种局限，也是一种资源。它限制了获取中国治理中细微、日常质感的渠道——那些不成文的规范、非正式的权力结构，以及任何正式框架都无法完全捕捉的现实生活。但它也赋予了诊断的自由，这种自由是接近权力中心的人往往被劝阻的；它也赋予了诚实陈述结论的意愿，而这种意愿往往受到制度立场的约束。本报告不自称拥有内部消息，而是主张拥有一种连贯的视角——这种视角可能对那些身处制度位置并正在寻找能解释其所经所感的框架的人有所帮助。

作者注意到，从中国境外撰写关于中国的报告具有特殊的困难。中国已成为一个高度极化的讨论话题，大多数外部评论倾向于非黑即白的分类：要么是单纯的赞美，要么是彻底的谴责。本系列报告试图避开这种二元对立，将中国视为一个正面临极高风险挑战的极其复杂的治理案例，并运用其在分析成熟民主国家失败时所使用的同样严谨且不带偏见的标准进行诊断。

关于方法论：治理即工程

本报告中使用的分析逻辑基于**“治理即工程” (Governance as Engineering)** 这一方法论。这一方法论并不将国家视为一个法律或意识形态实体，而是将其视为一个**反馈控制系统**。

其核心关注点在于：

1. **现实模型**：治理系统用来理解世界的现实模型是什么？
2. **执行能力**：系统将指令转化为行动的能力如何？
3. **校准反馈**：系统如何检测其行动何时偏离了现实，以及这种检测的准确度和速度如何？

本报告对中国“校准赤字”的诊断并非政治评判，而是一个系统性结论。如果一个反馈回路被延迟、扭曲或压制，那么无论其执行机器多么强大，系统最终都会与其治理的领土发生脱节。

全系列概览

本报告是中国治理研究的结语，也是“治理赤字”全系列研究的第十一篇报告。该系列旨在绘制 21 世纪全球治理失败的地图：

1. **德国**：侧重于“执行赤字”。
2. **法国**：侧重于“整合赤字”。
3. **瑞典**：侧重于“反馈赤字”。
4. **印度**：侧重于“同步赤字”。
5. **欧盟**：侧重于“连贯性赤字”。
6. **英国**：侧重于“控制与交付错位”。
7. **巴西**：侧重于“积累赤字”。
8. **俄罗斯**：侧重于“权力纵向赤字”。
9. **美国**：侧重于“整合赤字”。
10. **芬兰**：侧重于“吞吐量约束”。
11. **中国**：侧重于“校准赤字”。

这十一份报告共同构成了一个全球诊断框架，涵盖了从第一层级的执行与整合失败，到第二层级的速度与吞吐量约束，再到威权熵增与威权过度整合的边界条件，以及高信任民主国家在解决首要挑战后发现的次要限制。该系列并不自封完备，而是自视为一个基础，供进一步分析、深度测试和更好的制度设计参考——也是对该框架自身局限性的诚实承认。