

AGI 时代的系统科学与复杂性科学重新对账

——涌现、跨尺度反馈与文明治理的推导边界

Reckoning with Systems and Complexity Sciences in the Age of AGI

Emergence, Cross-Scale Feedback, and the Inferential Boundaries of Civilizational Governance

作者 子君赋 出品 文明跃迁研究组

独立发布版 V1.3 维度跃迁与跨学科因果关系修订

面向系统科学、复杂性科学、网络科学、复杂适应系统、社会—生态系统、系统安全及跨学科治理研究共同审查；独立分析边界、尺度、反馈、涌现、级联、反身性、局部优化与系统退居，并明确系统证据进入工程、政治、法律、伦理、经济与社会判断时的交接边界；不构成特定工程系统的安全保证、监管合规证明或治理参数手册。

正文修订 2026 年 10 月 10 日 作者 子君赋 出品 文明跃迁研究组

摘要

高级人工智能进入跨领域预测、资源调度、公共基础设施与多智能体协作后，系统科学面对的关键问题不再只是如何把更大系统看得更完整，而是模型、指标和干预本身开始改变被研究对象：边界选择影响谁被计入，预测改变行为与数据生成，局部优化经共享依赖形成跨尺度外部性，系统能力又可能被误写成主体性、合法性或主权。系统科学因此必须保持自己的专业独立：说明边界、尺度、反馈、级联与涌现如何成立；同时也必须标明证据在哪里止步，何处需要由工程、政治、法律、伦理与其他学科重新判断。局部判断可以分别成立，聚合、耦合与反馈后的整体后果仍可能失灵。本文提出“机制—尺度—反事实审计”，要求任何承担政策重量的涌现主张说明对象、基线、空间与时间尺度、生成机制、竞争解释、过程证据以及可区分的干预或反事实；同时建立跨层非转换原则：宏观图样不自动产生新实体，功能整合不自动证明体验，自组织秩序不自动获得授权，韧性不自动等于良善，临界信号不自动触发政治紧急权力，跨尺度协调需求也不自动生成单一中央主权。文章进一步提出边界四分法、双重反最优、反身证据更新、条件性模型多元、分布因果与责任交接，以及“系统退居”作为退出一分拆—转型的可审计制度提案；并以最小共享依赖机制例说明局部优化如何在组件均局部合理时生成系统损害。通过小世界与无标度网络之争、复杂性—稳定性研究、整湖早期预警实验及其误报边界、2003 年北美大停电等材料，本文区分已有较稳定支持、前沿竞争性解释、思辨性跨学科连接与公开规范承诺。文明跃迁仅提供最低哲学接口：智能、意识、主体、责任、权利与意义承担者之间不再能够凭历史重合互相推定；专业事实反过来有权修订该接口。本文的核心结论是：复杂性科学能够说明整体模式如何形成、何时改变及模型在哪些边界内失效，却不能单独说明整体是否具有体验、秩序是否正当、风险应由谁承担或权力应归谁。本文进一步将本学科职责前移至目标与目的生成：从利己为先转向利他为先的利己，使真实利害关系者在目的生成与锁定前进入判断。学科事实、方向、方法与结果分别说明，不以局部成败替代共同存续和三个长期目标。本文据此把研究从分清专业对象推进到深入跨学科关系，重点追踪本领域判断怎样经其他领域改变现实，又怎样由反馈返回并改变原条件与下一轮目的。多主体、多阶段和多场域的利益须分别呈现，相关、共同原因、中介与因果反馈须分别检验，新增关系维度接受具体反证。

关键词：高级人工智能；系统科学；复杂性科学；涌现；多尺度反馈；反身性；临界转变；韧性；系统退居；文明治理

Abstract

As advanced artificial intelligence enters cross-domain prediction, resource allocation, public infrastructure, and multi-agent coordination, the central systems question is no longer merely how to observe a larger whole. Models, metrics, and interventions increasingly alter the systems they describe: boundary choices determine who is counted; predictions reshape behavior and data generation; local optimization propagates externalities through shared dependencies; and system-level capability may be mistaken for subjecthood, legitimacy, or sovereignty. Systems and complexity sciences must therefore remain professionally independent in explaining boundaries, scales, feedback, cascades, and emergence, while also marking where their evidence ends and where engineering, political, legal, ethical, and other judgments must take over. Locally valid conclusions can still combine into system-level failure. This paper advances a mechanism-scale-counterfactual audit for emergence claims that carry policy weight. Such claims must specify the object, baseline, spatial and temporal scales, generative mechanism, competing explanations, process evidence, and discriminating interventions or counterfactuals. It also proposes a cross-level non-conversion principle: macro-patterns do not automatically establish new entities; functional integration does not establish experience; self-organized order does not authorize itself; resilience is not goodness; dynamical warning signals do not create political emergency powers; and cross-scale coordination needs do not generate unitary sovereignty. The paper develops a fourfold boundary analysis, a double rejection of both local and single-global optimization, reflexive evidence updating, conditional model pluralism, a handoff from distributed causation to responsibility analysis, and system retreat as an auditable proposal for exit, decomposition, and transformation. A minimal shared-dependency example shows how locally rational optimization can generate system damage even when no component is obviously wrong. Evidence from network science, complexity-stability research, whole-lake early-warning experiments and their false-

positive limits, and the 2003 North American blackout is separated into established support, competing frontier interpretations, speculative interdisciplinary connections, and explicit normative commitments. Civilizational Transition supplies only a minimal philosophical interface: intelligence, consciousness, subjecthood, responsibility, rights, and meaning-bearing can no longer be inferred from their historical co-instantiation in human beings. Scientific evidence may in turn revise that interface. Complexity science can explain how system-level patterns form, change, and fail within declared boundaries; it cannot by itself determine whether a whole experiences, whether an order is legitimate, who ought to bear risk, or where public power belongs. The paper further locates disciplinary responsibility before goal and purpose generation: a shift from self-interest first to self-interest with others first brings subjects with genuine stakes into judgment before goals are formed and locked. Facts, direction, methods, and outcomes remain distinct. The analysis is further extended to cross-disciplinary causal relations. It traces how judgments in this field change real-world conditions through other domains, and how feedback modifies the original conditions and subsequent purposes. Effects across actors, stages, and settings must be distinguished from correlations, common causes, and merely asserted connections.

Keywords: advanced AI; systems science; complexity science; emergence; cross-scale feedback; reflexivity; critical transitions; resilience; system retreat; civilizational governance

研究定位、材料范围与证据纪律

本文的研究起点是理论与维度跃迁。保留本学科的专业精度，同时深化它与相邻领域的因果关系；专业边界标明证明责任，不终止对共同后果的追踪。既有学术参与竞争解释和反例检验，不能以原分类尚无位置要求新问题退回旧框架。

本文定位为概念性系统分析、复杂性治理方法论与跨学科规范接口研究，不是一般系统论史、系统综述、具体系统动力学模型、工程安全认证或现行法意见。材料以截至 2026 年 8 月 24 日可核验的原始论文、出版社文本、官方事故报告与同行评议研究为主。经典理论用于恢复概念的原始适用条件；案例用于识别机制和外推边界；任何传统与本文自身都不享有免证地位。

全文将主张分为四类论证与证据身份，为便于核验简称 A/B/C/F。A 类是已有较稳定支持的结构、机制或限定事实；B 类是存在模型依赖、领域差异或竞争解释的前沿判断；C 类是尚未由现实对象充分检验的跨学科连接与 AGI 情景；F 类是公开署名的哲学/规范承诺，例如意义来源保持多元、系统复杂性不能自动生成权利或主权、真实主体与受影响者不能被总体指标静默抹除。A 类结论仍须绑定对象、尺度、时间窗和模型假设；B 类必须同时呈现反例与外部效度；C 类只承担压力测试，不能被写成已经发生的科学事实；F 类不伪装成经验真值，而以价值前提、逻辑一致性、反例与现实后果接受审查。经验、解释、规范判断和制度建议分别举证。

陌生读者入口：文明跃迁的最低接口

读者无须预先了解或赞同文明跃迁即可独立审查本文。本文只采用三项可修订前提。第一，意义是实际主体在生活、关系、实践、选择与承诺中形成的多元价值维度，不是系统状态、持久性指标或统一目标函数，意义也不是公共秩序的唯一价值。第二，有限性至少包括知识、行动时间、物质资源、生命历程与现实修复能力的有限；它使时滞、分配和不可逆后果具有重量，却不能由复杂性科学证明意义的唯一来源或任何具体制度义务。第三，系统是针对问题、尺度和时间窗划定的分析单位，主体则需要体验、连续性及自身后果关系等独立证据支持，二者不共延。

在此接口下，AGI 时代不是对某一能力阈值或已完成机器主体的断言，而是指高级人工系统使智能、意识、主体、责任、权利、政治资格与意义承担不再能够凭其在人类身上的历史重合而互相推定。当前公开证据仍不足

以证明当前高级 AI 具有主观体验、不可转移的主体连续性和自身后果意识。在责任可追溯、明确授权、可审计、可纠错、可退出的具体部署中，当前高级 AI 可以承担意义理解、保护、传递与能力外延等功能角色；这描述的是可问责部署中的角色，不是系统本性，也不构成主体性、资格或权力证据，同时保持未来主体证据开放。系统科学提供专业事实并有权迫使上游哲学前提重开，哲学只规定非自动推导边界，不裁决涌现、反馈、临界转变或韧性是否存在。

论文的最小承诺

本文不要求读者接受一套封闭的文明理论。只要接受“系统边界需要说明、跨层推导需要桥接、专业事实可以修订哲学接口”三项原则，其余命题均可在系统与复杂性科学内部独立检验。

目录

摘要.....2

Abstract.....2

研究定位、材料范围与证据纪律.....3

系统研究的多重边界与目的内生化.....6

从风险反馈到目的递归.....7

一、 导论：从更强系统到可治理复杂性.....7

二、 理论遗产与未清债务：哪些工具仍然有效.....8

三、 对象、边界、尺度与观察者：系统定义本身需要负责.....10

四、 涌现的机制—尺度—反事实审计.....11

五、 跨尺度反馈、时滞与反身证据更新.....12

六、 双重反最优：局部理性与单一总体目标的共同限度.....13

七、 自组织、秩序与合法性：形成机制不替正当性作答.....14

八、 路径依赖、临界转变、韧性与系统退居.....15

九、 多智能体、系统之系统与分布因果.....16

十、 现实材料与构造压力测试：机制不能由案例名代替.....17

十一、最强反驳、败北条件与开放研究议程.....18

结论：整体能力越强，推导边界越需清楚.....20

续读与共同研究入口.....21

附录 A 来源—主张边界登记.....25

附录 B 证据、外推与跨层审计清单.....27

参考文献.....28

系统研究的多重边界与目的内生性

整体认识必须能够重开边界

系统科学已经深入整体、反馈、尺度和涌现。AGI 条件下的新任务，是把观察者、模型、制度与目的生成纳入同一可修订研究过程。边界不是一次选定的世界边缘：一个领域的优化若在边界外制造成本，研究必须重开边界，又不能因无限扩展而使机制无法辨认。

从结果反馈进入目的递归

假设调度系统提高单个网络的效率，却增加其他网络的共用负荷与依赖。压力改变组织的安全需要，安全需要推动冗余或集中，结构变化又改变下一轮调度目标。工程、经济、政治和社会研究共同辨认这条链；系统研究说明时滞、阈值、级联与反身性，而不是把这些概念作为规范结论的证明。目的在链中变化，不能始终被设为模型外部常数。

多尺度结果不能相互抵销

实例、组织、基础设施、地区与代际尺度可能呈现不同成效；启动阶段快速集中可能解决迫近风险，也可能留下稳定阶段难以退出的依赖。均值改善不能抵销某个关键节点的不可逆封闭。多个主体的模型应保留利益冲突、拒绝与不合作条件，不能预设完美协调再声称整体已经成立。

怎样检验理论增量

检验加入目的更新、跨域依赖和慢变量后，是否得到不同且更可核查的解释或风险预警。反馈机制应有独立过程证据，不能仅由曲线共变宣布。若更简单模型同样解释结果，或新增关系只增加不可检验参数，应削减模型；若遗漏后果不断返回并击败预测，则须修改系统边界和原先的局部结论。

从风险反馈到目的递归

系统科学应将目标生成也作为系统的内部变化。新能力改变关系，新关系改变安全与利益需要，需要又进入下一轮目的；若只固定一个外部目标优化行动，便遗漏了文明递归最重要的变量。利己为先与利他为先的利己，在此产生不同默认路径。

旧路径可以表示为博弈、合作控制博弈、新利益、再博弈；新路径则使真实利害在目的形成时进入，令合作、新需要与新合作相互生成。冲突仍然存在，但合作能否成为生成逻辑，应从下一轮目的的实际变化判断，而不是从一次行动的善意宣言判断。

人类之间与人类—AGI 之间的递归必须分别标识行动者、能力变化、控制关系和不可逆后果。前一链不需要机器意识；后一链须说明更强能力与自主目的的适用条件。两链耦合时要追踪谁借另一关系增强自身权力，不能以一个抽象整体代替具体主体。

形式模型可以展示某些机制，不能因参数设定而替全部根判断取得证明。比较应公开边界、更新规则、可观察变量与竞争模型，检查无外部回报、单点失效、恶意行动者及关系权重寻租。模型中的合作若来自预设完美善意，就没有检验生成机制。

局部成效与全局后果须并列，安全余量和纠错空间不能被平均收益掩盖。反证应指出某项递归如何长期被阻断，或哪种更强解释给出不同因果去向；一次模拟获胜、局部风险下降或一段和平都不能单独结算共同未来。

一、导论：从更强系统到可治理复杂性

系统科学长期处理整体、关系、层级、控制与环境之间的结构，复杂性科学则研究非线性互动如何产生难以由局部直觉直接推出的宏观模式。AGI 没有使这些问题第一次出现，却改变了它们的制度密度。一个跨领域模型不仅观察系统，还可能分类对象、配置资源、修改接口、调用其他代理并执行行动；多个系统又共享云、身份、支付、能源和通信基础设施。于是，预测结果能够改变被预测者，局部成功能够通过依赖网络转移损失，部署本身能够重写下一轮证据。

问题由此从“怎样建立一个更准确的整体模型”转向“怎样在模型不完备、主体异质、反馈反身、损失分布不均且部分后果不可修复时，使复杂性保持可观察、可争议、可停止与可修订”。本文不为此另造一个凌驾于学科之上的“整体科学”。这里只讨论复杂系统进入文明尺度后的治理条件：当技术、制度、生态与主体持续共同演化，观察和干预进入因果链，后果跨越组织、法域与代际，且系统变化可能影响存续、权利或不可补偿的生命史时，工程性能不再穷尽问题。

这里的“文明尺度”不把文明拟人化，也不预设一个统一意志。它只标记影响的尺度、持续性与跨领域耦合；“可治理”也不等于可被单一中心控制，而是指边界、假设、分歧、权力、责任和退出条件能够进入公共审查。文明可以形成整体能力，却不能因此被写成一个具有第一人称体验、代表关系或主权的超主体。

1.1 三项核心贡献

第一，本文把“涌现”从泛化修辞还原为可审计主张，提出对象—基线—尺度—机制—竞争解释—过程证据—反事实的七项审计。它既不要求宏观解释被微观方程完全演绎，也不允许“看起来新”“难以预测”或“一张模拟图”直接承担实体、主体或政策结论。

第二，本文提出跨层非转换原则。系统科学结论可以为规范与制度研究提供事实前提，却不能越过独立论证：自组织不是自授权，稳定和韧性不是良善，因果中心性不是法律责任，预测准确不是目标设定权，协调需要不是主权，功能统一也不是主体性证据的充分条件。反向同样成立：意义多元、有限性与主体开放性不能证明某系统存在涌现、逼近临界点或必须采用冗余、多中心与退居。

第三，本文把边界、尺度与观察者从纯模型选择扩展为可问责对象，并提出双重反最优：局部行动者的理性会产生总体外部性，单一“总体最优”也可能抹去分布、权利、地方知识和不可通约损失。对应的制度增量不是一个复杂性总分，而是让对象边界可重开、模型假设可比较、慢变量可持续观察、责任可交接、系统可退出或转型。

1.2 研究问题与边界

表 1 本文的五个研究问题与不承担的任务

研究问题	本文承担	本文不承担
涌现如何成为可检验主张？	给出机制、尺度、竞争解释和反事实审计。	裁定所有强涌现或本体论争论。
反馈如何改变证据？	区分部署前后对象、数据生成与反身影响。	声称所有社会预测必然自我实现。
局部与整体如何对账？	分别展示总量、分布、外部性和不可交易底线。	提供唯一文明目标函数。
系统能力如何进入治理？	阻断从能力到主体、授权和主权的自动推导。	替政治学和法学确认具体权源。
高不确定性下怎样行动？	提出风险相称的模型多元、可逆试验与退出设计。	用不确定性阻止所有部署或量化。

1.3 系统科学能够独立回答什么，到哪里必须交接

系统科学与复杂性科学能够独立回答的是：在明确对象、环境、尺度与时间窗以后，边界如何改变结论，反馈、时滞、耦合、网络结构和局部互动怎样生成总体模式，组件或子系统分别正常时为何仍可能出现系统级失效，以及哪些干预与反事实能够区分竞争机制。它也能够比较不同结构在级联、恢复、锁定、冗余、适应和退出上的条件性后果。

它必须停止在跨学科推导的交接点：从“系统存在某种整体模式”到“谁具有体验、谁拥有权利、何种损失不可牺牲、谁有权授权、谁应承担责任”，都需要认识论、生命与心灵研究、伦理、法学、政治学、经济学和社会研究重新进入。系统层的“整体”不是上位裁决者。专业边界在这里既保护机制判断的精确，也标明证据在哪里止步、下一判断由谁接上；局部判断即使分别成立，也不能自动证明聚合、反馈和冲突后的整体方向正确。

二、理论遗产与未清债务：哪些工具仍然有效

一般系统论把开放系统、环境、层级和关系带入跨学科分析；控制论建立反馈、调节与通信的共同语言；西蒙的近可分解性解释了为何层级与模块有助于复杂结构的形成与分析；安德森的“More Is Different”强调尺度变化可能要求新的组织概念；Conant—Ashby 定理则在特定假设下连接了成功调节与对象模型。[1-5,22] 这些工作仍是有效工具，却没有替后来的研究回答目标由谁设定、边界如何选择、系统输出怎样进入公共权力。拥有更好的系统模型，也不等于拥有完整现实或统治资格。

非平衡自组织、自组织临界、复杂适应系统与网络科学展示了无逐步中央指令的有序结构、局部规则与宏观模式、小世界结构、增长与优先连接等机制。[6-10] 但跨域借用常把机制压成隐喻：某个分布看似幂律，便宣布

系统“自组织临界”；某个网络存在枢纽，便宣布集中不可避免；某种秩序自发出现，便推定它更自由或更正当。Broido 与 Clauset 对近千个网络的检验显示，强无标度结构并不普遍；这不是否定幂律，而是提醒模型形式必须与替代分布竞争。[11]

复杂性—稳定性研究同样反对口号化。May 在随机相互作用模型中表明规模、连接度与相互作用强度可能改变稳定条件；后续研究进一步显示相互作用符号和网络结构会改变结论。[12–13] 因而“复杂越稳定”与“复杂越脆弱”都不能脱离对象结构成为普遍定律。多样性、冗余、模块化和集中在不同耦合条件下可以分别提高或降低韧性。

生态韧性、社会—生态系统、临界转变与早期预警研究，为多稳态、适应、转型和跨尺度治理提供了重要框架。[14–21,33–34] 这些研究内部已经包含边界意识：韧性必须问“何物对何种扰动维持何种功能”，预警信号需要机制、对照和误报分析，多中心安排也只有在特定规则、能力和外部性条件下成立。本文的增量不是纠正这些成熟文献，而是阻止它们在 AGI 制度桥接中被压缩成“越韧越好”“出现信号就扩权”“多中心天然民主”或“全局模型天然优越”。

系统安全遗产进一步说明，整体故障不能还原为单组件错误。Perrow 的“正常事故”传统强调复杂交互与紧耦合下的意外组合，Leveson 则把安全分析推进到控制结构、约束与组织过程。[30,49] 本文相对这些遗产的新增量，不是再次证明“复杂系统会出事故”，而是把两种 AGI 时代条件纳入同一审计：第一，模型、指标和部署会反过来改变被研究对象与下一轮证据；第二，系统能力进入公共资源与权力链后，系统保证的证据必须在授权、权利与合法性处显式交接。前者要求证据版本化，后者要求系统层的“整体”不能吞并政治、法律和伦理的独立判断。

2.1 科学命题、接口命题与制度提案分开

表 2 三类命题的证据所有权

命题层	典型问题	可使用的证据	不得冒充
复杂性科学／方法	边界敏感、涌现机制、反馈、时滞、级联、模型失效。	数据、模型、过程追踪、干预、反事实、外部验证。	价值排序、法律效力或公共授权。
规范与政治接口	形成是否等于正当；协调是否需要集中；损失如何进入决定。	事实前提加上公开的权利、非支配、分配与授权论证。	复杂性科学定律。
本体论交接	功能统一是否构成主体候选；系统边界是否等于主体边界。	心灵、生命、认识论及制度程序的独立联合证据。	由网络图样直接确认或排除主体。
制度提案	退出、分拆、监测、模型竞争和申诉如何设计。	比较实施、失败率、行为后果、成本、权利与法源。	从机制直接推出的唯一方案。

2.2 证据与论证身份 A／B／C／F

A 类结论包括：系统结论对对象和边界敏感；反馈、时滞、网络依赖与局部互动可产生非直觉总体后果；特定受控环境中可观察到候选预警；指标使用在部分社会情境中会改变行为；现实电网事故能够呈现组织—技术共同级联。[9–10,12–25,31–34,37–41] B 类包括：宏观尺度是否具有更强因果效力、现实网络是否严格无标度、某些社会系统能否被稳定识别为临界转变、不同治理结构在跨境级联中的相对优势，以及多智能体生态会出现何种新群体行为。[11,26–30,39–46] C 类包括：未来机器群体主体性、文明级单一临界点、AGI 自动形成公共意志或

新主权。F 类则承载本文公开的哲学／规范前提与桥接禁令，例如复杂性不能自动生成权利、授权、人格或主权，意义与受影响主体不能被统一指标静默结算。本文不把 B 或 C 写成已证实科学，也不把 F 写成复杂性科学的经验定律。

三、对象、边界、尺度与观察者：系统定义本身需要负责

系统不是自然世界自动附带的唯一切割。研究者基于问题选择对象、环境、变量、尺度和观察窗口；这种选择可以合理，却必须可见。城市交通优化若以行政区为边界，可能看不到通勤、能源和住房后果；平台推荐若以点击窗口为边界，可能看不到长期信任与公共讨论；公共基础设施若按供应商品品牌计算冗余，可能遗漏共同云、身份或能源故障域。边界不完备不是停止研究的理由，却是登记遗漏、做敏感性分析与设定重开条件的理由。

3.1 边界四分法

表 3 四类边界不能相互替代

边界类型	核心问题	典型证据	常见越界
因果边界	哪些变量与关系显著影响目标结果？	领域数据、机制、干预与时序。	把相关性或邻近性当成完整因果域。
模型／观察边界	哪些对象、变量、时间窗与尺度被表示？	测量协议、抽样、模型假设、遗漏分析。	把模型之外写成现实之外。
责任边界	谁设计、授权、控制、受益、预见、监测、停止和修复？	组织角色、权限、日志、能力与行为。	由因果中心性直接分配过错。
法律／管辖边界	何种法源、法域、权利与救济适用？	法律文本、机构权限与司法判断。	由系统图自动生成法域或授权。

被模型纳入同一系统，不等于形成同一主体；被排除在模型边界之外，也不取消其受影响地位。分析边界、道德共同体边界、法律资格边界和政治管辖边界必须分别说明。一个流域可以是恰当生态分析单位，却不是一个自动拥有统一偏好或代表关系的政治主体；一个平台生态可以构成功能网络，却不因此取得用户的共同意志。

3.2 五个尺度与跨尺度非等价

个体或节点、关系或网络、组织或制度、系统之系统、文明与代际，是本文使用的五个分析尺度。宏观模式可以是真实且有解释力的，却不能把微观主体的体验、权利和损失吸收为平均值；微观因果也不能穷尽宏观结构的约束。跨尺度分析应同时回答：新变量在何种聚合规则下出现，向下或向上的干预如何改变结果，结论是否对尺度和时间窗稳健，以及责任能否从因果描述中被独立交接。[48]

尺度变化不只增加细节，也可能改变问题。短期恢复速度与长期替代能力、节点可靠性与网络级联、个人代理忠实与公共资源拥塞、城市效率与流域生态，不是同一指标在不同分辨率下的简单版本。所谓“整体更好”必须说明整体指谁、时间窗多长、何种损失被平均、哪些主体不能退出。

3.3 观察者何时进入系统

“观察者是系统的一部分”不是所有系统研究的普遍定律。更精确的条件是：当模型公开、分类改变资源、指标被用于奖惩、被观察者能够策略响应，或预测触发干预时，观察与部署可能进入因果链。Goodhart、

Campbell 以及关于排名反身性的研究表明，指标一旦成为高后果目标，关系可能失稳、行为可能被重构；但这并不意味着任何量化都会失败。[23–25] 研究者需要比较部署前后数据生成过程，并把模型带来的行为改变视为新对象。

四、涌现的机制—尺度—反事实审计

本文把涌现定义为：相对于明确基线和尺度，在要素互动中出现的非平凡宏观模式、能力或解释变量。稳定性不是定义前提，而是待检属性；暂态图样同样可能具有解释意义。“规范涌现”则不作为科学类型使用，因为规则事实上形成与规则取得正当性属于不同问题。

4.1 七项最低审计

表 4 承担政策重量的涌现主张须回答的七个问题

审计项	必须说明	失败信号
对象	要素、关系、环境与状态变量是什么？	只说“系统整体”而无对象。
基线	与简单聚合、共同外因或随机图样相比新在何处？	新奇只来自展示方式或阈值选择。
尺度	空间、组织、时间和观察尺度是什么？	换尺度后结论消失却未说明。
生成机制	局部规则、网络结构、约束、资源与反馈如何产生图样？	以“复杂所以涌现”循环解释。
竞争解释	共同驱动、测量伪象、选择偏差和替代模型有哪些？	只有一张模拟图或单一拟合。
过程证据	时序、传播、介质、阈值与中间状态是否可观察？	只观察起点与终点。
反事实／干预	改变哪项条件会产生可区分结果？	任何结果都被解释为涌现。

这套审计不否定宏观解释的相对自主性。Bedau 的弱涌现、关于宏观因果效力的研究及其批评，都说明“宏观是否胜过微观”依赖计算、干预分布与度量选择。[26–28] 本文只要求：若宏观性质被用来配置公共资源、限制权利或扩大权限，它必须与生成条件和可区分的后果连接。一个原则上拒绝所有反事实区别的形而上主张可以继续讨论，却不能直接成为经验政策证据。

4.2 临界、级联与不可逆必须分别定义

“临界”至少可能指物理相变、自组织临界、临界减速、网络阈值或一般政策转折。每次使用都应说明动力系统类别、状态变量、候选机制、稳定域或吸引域、控制参数与网络条件；突然变化不自动等于临界转变。本文所称级联，是初始扰动通过依赖、传播或负载转移产生连续次生变化，必须给出拓扑、传播规则、时间顺序和反事实。相继发生不等于因果级联。

不可逆也不是单一属性。状态映射不可逆、动力学迟滞或路径不可回返、现实时间与成本下难恢复、以及生命、关系和历史上的不可补偿，具有不同证据与学科所有权。系统科学可以检验状态转移、迟滞和恢复路径；某项损失是否不可补偿，则须由生命科学、伦理、法律及受影响主体另行论证。任何动力学不可逆都不得直接推出道德绝对性或政治紧急权力。

4.3 推导硬停止

涌现硬停止

若主张只有单一模拟或图像、未声明尺度与基线、未比较替代解释、对象因部署而改变却仍沿用旧证据，或从功能图样直接推出意识、人格、合法性与主权，就不能继续承担高后果制度推导。

五、跨尺度反馈、时滞与反身证据更新

反馈不只是正反馈和负反馈的教科书分类。文明级系统同时包含增强、平衡、跨域、反身与共同演化反馈：资源分配改变行为，行为改变数据，数据重训模型，模型又改变下一轮资源；基础设施效率提高使用量，使用量又扩大共同依赖；风险标签带来支持，也可能造成污名与数据回避。动态决策实验也表明，时滞与累积会使行动者误判自身行动的回流。[47] 对这类对象，证据不是部署前收集一次便永久有效。

5.1 快变量、慢变量与观察窗口

交易成功率、响应时间和局部故障可以按分钟观察；技能流失、制度能力、社会信任、生态恢复与地方知识则可能按年或代际变化。只用快变量评价系统，会把慢性锁定误写为即时成功。相反，只等待所有慢变量完全显现也会造成行动瘫痪。合理做法是双窗口登记：明确短期性能、慢变量候选、可观察代理、复核周期与停止条件，不把一个窗口的准确率外推为系统生命周期的保证。

5.2 延迟、迟滞与修复不是同一件事

延迟是输入变化与可观察输出之间的时间间隔；迟滞则意味着回程路径不同，停止输入也不能立即返回原状态。系统停止、配置回滚、功能恢复和现实修复必须分别验证。一个模型版本可以回滚，已经改变的资源机会、关系信任、地方能力和生命事件却未必回到原点。技术可逆性因此不能充当现实可修复性的替代指标。

5.3 证据版本化

表 5 反身系统中的证据更新程序

时点	最低问题	允许结论	重开触发
部署前	对象、边界、基线、数据生成、竞争模型是什么？	在声明环境中的候选机制或性能。	对象、接口、规模或用途变化。
试点中	观察或干预是否改变行为、选择和数据？	限域因果与副作用判断。	策略适应、分布偏移、权限扩散。
扩散时	新地区、群体、供应链与法域是否仍属同一对象？	新对象需重新验证，不继承全部试点资格。	依赖集中、外部性跨域、替代减少。
长期运行	慢变量、时滞、锁定、退出与责任尾期如何变化？	持续但可修订的制度判断。	预警、申诉、共同故障或主体证据变化。

模型多元不是所有问题的永久仪式。低后果、可逆、边界稳定且主模型经异质环境长期检验占优的问题，可以采用单一主模型并保留失效监测；结构不确定、模型选择显著改变分布结果、后果不可逆或跨尺度外部性较大的问题，才提高竞争模型、情景与敏感性要求。模型多元不能无限延期决定，单模型速度也不能购买盲区。

六、双重反最优：局部理性与单一总体目标的共同限度

复杂系统的第一类失效来自局部理性的聚合。行动者各自在既定规则内优化价格、注意力、速度、预约、收益或安全，整体却可能发生资源耗竭、拥塞、极化、军备竞赛与级联。Schelling 的隔离模型表明温和的局部偏好可以产生强烈宏观分离，却不证明现实隔离只有一个原因；Hardin 展示特定开放获取条件下的公地困境，Ostrom 则以长期制度研究反对“私有化或中央强制是唯一解”的万能药。[31–34] 宏观图样不能直接反推每个行动者的动机，局部守规也不能证明总体后果可接受。

第二类失效来自单一总体最优。全局模型能够降低某种成本，却可能在目标函数中删除分配、基本权利、地方知识、关系连续性与不可补偿的生命史。平均改善并不说明谁承担尾部风险；总体稳定也不说明被维持的状态值得保存。AGI 扩大了搜索和协调能力，却没有消除目标设定的政治性，也没有把不可通约价值转换为同一单位。

6.1 总量与分布必须双账

任何高后果优化至少同时展示：总体性能、群体与地域分布、不可退出者、尾部风险、跨期转移、关键权利侧约束、不可补偿损失，以及替代方案。分布审查不是否定总体指标，而是阻止总体指标取得牺牲具体主体的默认许可。一个方案可以在总量上更优，仍因把重大伤害集中于无申诉能力者而不可接受；一个地方也不能以自身知识与自治为由把伤害转嫁给外部群体。

6.2 不可通约不等于无法决定

价值不可通约不要求制度停摆。可以先设定不可交易底线，再进行多准则比较；对证据不足且后果可逆的问题采用分阶段试验；对不可逆后果提高举证与授权要求；保留异议、替代与复核；明确何种权力可以在何种时间窗内决定。拒绝一个文明总分，不等于拒绝比较，而是把比较的维度、权衡与损失承担者显影。

6.3 指标反身性与累积劣势

当贡献、风险、信誉或生产率指标进入分配，指标会改变行动者可获得的下一轮资源。易量化、短周期和公开行为可能被高估，照护、长期基础工作与不可见贡献被低估；低记录者因资源减少而更难产生可见记录，既有分布又被系统解释为能力差异。该机制是构造性风险而非已证实的普遍事实。最低边界是把基本保障、人格、权利与政治声音从单一绩效分中分离，记录归因不确定性并允许异议与更正。

6.4 最小机制例：局部优化如何经共享依赖生成系统损害

为避免“双重反最优”停在概念层，下面给出一个明确标注的构造例，不把它当作现实频率或普遍定律。两项公共 AI 服务 A 与 B 都调用同一身份／云服务 S，分别以响应时间和本服务成功率为局部目标；控制器在本地拥堵时自动增加并发与重试。只要 S 的总负载处于安全区，每个局部策略都可能改善自己的指标；当 A 与 B 同时扩张，重试又共同占用 S，系统总负载接近共享容量，排队和失败概率可能非线性上升，重试进一步放大负载，最终两项服务同时下降。没有任何一个局部控制器必须“明显犯错”，系统损害来自共享依赖、时序与被局部目标遗漏的外部性。

表 6A 共享依赖下局部优化的最小构造例

局部动作	局部可见结果与共享效应	可区分测试
A 扩大并发	A 等待下降、吞吐提高；同时消耗 S 的公共容量、身份及网络资源	固定 B，仅改变 A 并发； 观察 S 余量与 A/B 尾延迟
B 同步扩张	B 局部指标改善；与 A 叠加后逼近共同容量边界	交换时序或隔离资源； 比较共同失效是否消失
失败后重试	单请求成功率可能短时上升；重试正反馈继续抬高 S 负载	限制重试或加入退避； 观察级联变化
系统级结果	两个局部策略都可“合理”；共享容量、时滞与反馈使整体损失上升	把共享风险写入局部目标或拆分依赖；若结果不变则 降级机制解释

可用一个最小形式表达这一条件性机制：局部控制器分别优化 $U_A(x_A)$ 与 $U_B(x_B)$ ，而系统损失还包含共享风险 $R(x_A+x_B)$ 。若局部目标没有计入 R ，且 R 在总负载接近容量 C 时陡升，则局部最优 x_A^* 、 x_B^* 的组合可能显著提高系统损失。关键不在这个简式本身，而在它生成可区分预测：拆分共享依赖、限制重试、把公共容量成本写入局部目标或改变负载时序后，级联风险应发生可观察变化；若这些干预不能改变结果，这一机制解释就应降级或退出。

七、自组织、秩序与合法性：形成机制不替正当性作答

本文所称自组织，是系统在外约束、资源、接口与规则条件下，未由单一中心逐步指令而形成组织。它不是“无设计”“无中心”“无外部投入”或“自然自由”的同义词。协议、产权、平台接口、资源分布和法律秩序可以深刻塑造自组织的可行空间；一个没有单点指挥者的秩序仍可能产生事实支配。

形成、功能与正当必须分开。生成秩序回答模式怎样出现；功能秩序回答它能否协调、维持或解决问题；规范与政治判断则要回答同意、权利、规则改变、少数保护、分布、救济与替代。自组织秩序可能经独立程序取得合法性，却不能自我授权。这里的“自组织不自授权”是逻辑—政治接口，不是复杂性科学的新经验定律。

7.1 设计与涌现不是二选一

现实制度常同时包含设计边界与涌现实践。公共设计可以限定不可越过的权利、互操作、责任、申诉和退出条件，同时允许地方规则、专业惯例与协作方式在边界内形成。复杂性证据有时支持分散和多样，有时支持集中感知或统一协议；结构选择应由级联速度、知识分布、外部性、可逆性和问责能力共同决定。集中职能即使在经验上必要，也仍需单独证明强制权源、范围、期限与监督。

7.2 跨层推导防火墙

表 6 已有证据不得自动推出的结论

已有证据	不得自动推出
宏观图样或尺度上的新变量	新实体、本体独立、主体性或统一意志。
功能整合、记忆、自我修复、目标保持	体验、统一第一人称、人格、代表或公共主权。
自维持、适应、繁殖类比	生命、感觉、道德资格或权利。
自组织秩序、网络共识或协议一致	同意、自由、合法性、自授权或共同意志。
稳定、持续、高韧性或恢复速度	良善、正义、安全、应当保存或成功。
临界信号、级联风险或高尺度协调需求	政治紧急状态、无限扩权、单一中央或世界主权。
因果中心性、控制能力或预测准确	法律过错、目标设定权、统治资格或主权。
模型边界、总体最优或平均改善	道德共同体、法域边界或牺牲具体主体的许可。
技术回滚、系统存续或复杂度	现实后果已经修复、文明意义或意义权属。

防火墙也有反向方向。哲学上的主体开放性、意义多元和有限性，不推出系统存在涌现、接近临界点、具有自组织能力或必须采用某种架构；政治上珍视多中心与非支配，也不证明分散系统在特定级联中更有效。系统科学能够研究“整体”，不意味着“整体”取得了对法学、伦理、政治、生命与主体问题的上位裁决权。任何一层都不能用自己最强的语言替下一层完成独立论证。

八、路径依赖、临界转变、韧性与系统退居

历史顺序会改变当前选项，但路径依赖不等于绝对不可逆。Arthur 关于递增收益和技术竞争的模型说明，小事件在特定反馈条件下可能影响技术锁定；现实平台是否锁定，还需证明转换成本、网络依赖、接口控制、数据迁移障碍、技能流失或制度承诺已显著压缩可行替代。[35] 对 AGI 系统，试点成功不能只登记性能，还应登记依赖积累、替代能力与退出成本。规模、用途、法域或供应链变化时，扩散应被视为新对象。

8.1 四种“临界”不能混写

表 7 四种不可互换的临界或自组织机制

概念	原始机制	最低证据	不得互推
耗散结构	通量维持的远离平衡有序结构。	开放受驱系统、非线性和能质交换。	社会规范、SOC 或自由秩序。
自组织临界	特定慢驱动—耗散—阈值模型自发接近临界态。	模型规则、尺度分布及竞争机制。	幂律或 1/f 噪声单独确诊 SOC。
混沌边缘	随机布尔网络等构造模型的有序—混沌边界。	明确网络、参数与性能定义。	普遍生命、组织或治理最优点。

概念	原始机制	最低证据	不得互推
临界转变	接近某些局部分岔时的临界减速及状态域变化。	状态变量、候选分岔、对照、误报漏报。	所有突然变化都属该机制或都有预警。

Scheffer 等综述了临界减速等早期预警机制；Carpenter 等在受控整湖实验中观察到部分指标先于食物网转变；Boettiger 与 Hastings 则显示检测性能依赖数据、机制和选择过程，常用指标可能产生误报与漏报。[18–21] 因而预警可以触发调查、独立复核、缓冲和可逆准备，不能单独宣布政治紧急状态或自动扩大权限。若某一领域获得外部验证的敏感度、特异度和决策效用，预警可以进入限定操作触发，但权源仍须另证。

8.2 韧性、适应能力与转型能力分开

本文把扰动响应分为韧性、适应能力与转型能力：韧性指系统在变化中保持基本功能、结构、身份和反馈的能力；适应能力指行动者影响这种韧性的能力；转型能力指现有系统不可接受时创造根本不同系统的能力。三者可以相互作用，却不得压成一个总分。[14–17] 每次使用韧性都必须绑定对象、功能、扰动集合、时间窗和性能边界。

恢复得快不等于长期更安全，维持得久也不等于状态值得保存。一个高度韧性的压迫制度、脆弱的单一供应商入口、或快速恢复但持续侵蚀地方能力的平台，都可能在不同维度得到相反评价。复杂性科学能够说明何种结构维持了什么，不能单独回答被维持的状态是否良善、谁为恢复付费、谁有权要求转型。

8.3 系统退居：一种制度提案而非复杂性定律

本文提出“系统退居”，指系统在使命完成、权限终止、依赖风险上升或替代结构成熟时，能够退出核心位置、分拆、迁移或转型，而不把既有依赖变成自动续期理由。它不是复杂性科学已经证明的定律，而是接受比较实施和失败检验的制度提案。最低维度包括数据可迁、接口可替代、权限可失能、人工与公共能力可恢复、地方知识可保留、现实损害有修复和责任尾期。

退居不等于简单关闭。对生命支持和公共基础设施，突然退出可能比继续运行造成更大损害，因此应区分紧急停止、降级运行、替代接管、分阶段迁移和长期转型。若退居设计持续造成更大不可逆损害且不能降低锁定或任意支配，该提案必须重构。

九、多智能体、系统之系统与分布因果

多个 AI 代理各自忠实执行用户目标，不保证群体结果可接受。公共预约、交易、广告、算力与注意力资源可能出现高速抢占和拥塞；共享信号可能形成排斥或串谋；权限可能经转授权和工具链扩散；每个设计者只负责局部模块时，整体后果又可能被归为“生态行为”。这些是 C 类前瞻压力测试，现实频率和机制需要机器行为、网络实验、组织研究和现场数据共同检验。[29–30]

9.1 系统之系统没有自动共同意志

系统之系统中的子系统拥有不同使命、所有者、责任与法域。协议可以协调状态，不能生成公共目的；数据融合可以扩大态势感知，不能消灭主体和管辖边界；整体优化可以降低某类成本，不能证明整体利益已被正当代表。网络共识、整体记忆、目标保持与自我修复只能证明相应功能统一，既不证明也不排除未来主体性。若出现体验、连续性和自身后果意识的独立联合证据，应转交心灵、生命、认识论和制度程序审查，而不能由系统科学先验确认或否定。

9.2 分布因果不让责任蒸发

不存在单一充分原因，不等于不存在可识别的责任关系。复杂性分析负责识别因果贡献、控制位置、信息可得性、共同依赖与停止能力；伦理、政治与法学再依据角色、授权、可预见性、受益、专业义务、过错与救济规则配置责任。因果位置本身不自动生成、取消或分配法律与道德义务。

责任交接至少追问：谁设计目标与接口，谁批准部署和规模扩张，谁控制关键数据与凭证，谁能看到预警，谁能够停止或隔离，谁从规模获益，谁承担外部性，谁必须保存证据与完成修复。复杂系统可能没有全能控制者，却仍存在不对称塑造者；“无人完全控制”不能成为“无人需要负责”的同义词。

9.3 协调需求不生成复杂性主权

跨境级联、气候风险与共同基础设施确实可能要求高尺度协调，甚至在特定时间窗内要求集中感知和行动。需要证明的是具体职能、必要性、替代不足与退出条件，而不是由风险规模直接推出不可质疑的统一中心。嵌套安排可以在高尺度锁定最低不可逆边界、共同核验和危机通信，在国家与区域层保留法律责任和公共能力，在地方层保留在地知识与申诉，在个体层保护知情、拒绝、参与和基本权利。其经验优劣须比较，不能由多中心或集中任一理念预先决定。

十、现实材料与构造压力测试：机制不能由案例名代替

本节把现实材料与构造情景分开。现实案例只支持其对象和证据窗口内的有限结论；构造情景用于暴露框架必须追问的问题，不证明相应机制在现实中具有相同频率、强度或因果结构。一个案例有多个原因不等于无法分析，一个模型能再现实验图样也不等于穷尽现实事故。

10.1 2003 年美加大停电：实际级联不等于单一网络故事

2003 年 8 月 14 日的北美大停电影响约五千万人和 61,800 兆瓦负荷，是跨区域电力级联的现实材料。美加联合调查并未把事故归结为一个拓扑指标，而是识别了植被接触、线路跳闸、软件与告警失效、态势感知不足、操作协调和保护动作的相互作用。[37] 该案例稳定支持：局部故障可以在依赖、监测和组织失效共同作用下升级；它不支持：Buldyrev 的相互依赖随机网络模型已经解释全部事故，或中央化必然是唯一补救。

案例还说明责任与因果需要交接。级联结果不能压缩为单一“根因”，却可以识别植被管理、软件保证、运营监测、协调协议与制度责任。复杂性不是责任消失的理由，而是要求把控制位置、信息可得与停止能力画得更精确。2021 年德州及美国中南部寒潮则主要涉及严寒下发电设备、燃料供应与气—电依赖造成的供需危机和受控减载；它不能与 2003 年传输网级联混称为同一事故类型。[38]

10.2 复杂性—稳定性与金融网络：机制具有条件性反转

May 的随机矩阵结果提示复杂、连接与相互作用强度可能形成稳定性约束；Allesina 与 Tang 进一步展示相互作用符号和结构会改变稳定条件。[12–13] 在金融网络模型中，Acemoglu 等发现更密连接在小冲击下可以分散风险，在足够大冲击下又可能扩大传播，形成“稳健—脆弱”条件性；Iyer 与 Peydró 的印度银行自然实验则提供了窄域的同业暴露传播证据。[39–40] 这些材料支持的不是“连接越多越危险”或“越分散越安全”，而是同一结构在冲击规模、负载转移和恢复机制变化时可能反转。

10.3 早期预警：受控成功与检测边界同时成立

Carpenter 等在受控整湖实验中，通过逐步增加顶级捕食者使食物网失稳，并以相邻湖泊作为参照；若干统计指标在状态转变前出现变化。[19] 这是一项有机制、有对照的生态证据，不是文明预警器。Boettiger 与 Hastings 说明，即使在有利条件下，早期预警的检测能力也受数据长度、噪声、模型选择与条件抽样影响；随机转变可能产生“检察官谬误”式误报。[20-21] 因而预警的正确制度含义是提高调查与准备，而不是把统计信号转换为确定倒计时。

10.4 网络模型：构造结果与现实识别分开

Watts—Strogatz 模型说明高聚类与短平均路径可以并存；Barabási—Albert 模型说明增长和线性优先连接能够生成重尾度分布。[9-10] 观察到短路径或重尾，不能单独识别生成机制。Broido—Clauset 在其严格定义与检验门槛下发现强无标度网络罕见；Voitalov 等采用正则变化框架提出不同识别标准。[11] 争论的稳定教训不是选择一个阵营，而是先定义“无标度”、公开尾部估计与样本选择，并与对数正态等替代分布竞争。

10.5 八类构造压力测试

表 8 构造情景的机制问题与推导边界

情景	必须检验的机制	不得写成
流域—城市—能源—粮食耦合	行政边界是否转移生态与下游风险；季节与代际慢变量是否被遗漏。	城市效率证明流域整体可接受。
公共卫生—预测—行为—信任	分类是否改变就医、迁移、披露与数据质量。	预测准确自动授权强制或永久分类。
大量个人 AI 代理	局部忠实是否产生拥塞、抢占、串谋与无代理者排斥。	个人授权聚合为公共同意。
推荐—讨论—极化	同质聚集、奖励结构、事件与政治动员如何竞争解释。	单一算法造成全部极化。
云—支付—能源—通信	名义冗余是否共享身份、网络、能源与供应链故障域。	品牌数量等于独立路径。
高能力系统试点扩散	对象、地域、接口、采购与人才依赖是否已改变。	试点成功自动生成普遍资格。
灾后单一平台恢复	短期恢复是否侵蚀长期公共、地方与替代能力。	恢复速度等于综合韧性。
贡献或风险指标分配	指标如何改变资源、行为与下一轮可见记录。	记录等于人格、意义、权利或政治资格。

八个构造情景共同暴露了本框架必须追问的边界遗漏、反馈重写、共享依赖、路径锁定与责任跨层问题。它们不能证明现实机制普遍存在，也不能替代数据。其价值在于形成可证伪研究设计：明确对象、比较机制、寻找过程证据、观察条件性反转，并登记何种结果会让原假设降级或退出。

十一、最强反驳、败北条件与开放研究议程

一套跨学科复杂性框架最容易通过不断增加限定词而免疫反驳。本文因此把逻辑非蕴含、经验失败、适用范围修订与规范接口重开分开：逻辑非蕴含由反例与桥接结构检验；经验机制由数据、干预和外部效度检验；制度提案由比较实施、成本、权利与行为后果检验。

11.1 八项最强反驳

表 9 反驳、回应与真正败北条件

最强反驳	本文回应	迫使收缩或放弃的条件
这是把政治哲学包装成复杂性科学。	科学层只报告机制与失效；规范层以公开前提作条件推导。	接口不能产生新观察问题或机制区分，应从专业主体删除。
涌现不可还原，反事实审计会消解它。	不要求完全微观演绎，只要求宏观主张与生成条件及可区分后果连接。	审计在目标领域持续排除有解释力的宏观变量且无增量。
宏观主体性本来可能涌现。	系统功能既不证明也不排除主体；未来联合证据触发跨学科再审。	若框架把系统永久排除为主体候选，开放性主张失败。
复杂风险有时确实需要单一中心。	承认条件性集中职能；权源、期限、监督与退出另证。	若框架无论证据为何都偏好分散，应放弃该偏好。
模型多元与边界重开会造成瘫痪。	按后果、可逆性与结构不确定性分级；低风险可用主模型。	要求没有改善判断却持续造成重大延误，应降级。
韧性非良善只是专业常识。	承认不是新定律；增量在面向 AGI 基础设施的可审计接口。	接口不能改变任何案例诊断或设计问题，应并回既有框架。
分布式因果根本无法追责。	因果分析只交接控制、信息与贡献；责任由独立规范和法律规则配置。	责任图无法比普通组织图增加任何可审计关系，应撤销。
开放系统边界永远不完备。	只要求目的相关说明、遗漏登记、敏感性与重开条件，不要求全知。	新增边界长期不改变解释、预测或分布，则审查应停止扩张。

11.2 经验失败、范围修订与接口重开

- 所谓涌现机制不能在干预或反事实中区分共同外因、简单聚合与测量伪象，应降级为描述图样。
- 边界或尺度变化在目标领域长期不改变解释、预测或风险分布，相关审查要求应降低。
- 模型公开、分类和部署不改变行为或数据生成，反身性要求不应机械适用。
- 早期预警在特定领域获得外部验证的敏感度、特异度与决策效用，可进入限定操作触发，但仍不生成权源。
- 单一模型在异质环境、漂移和对抗条件下长期优于竞争模型，模型多元可降为失效备用。
- 路径效应经迁移实验和长期跟踪可忽略，不应继续声称锁定。
- 系统退居持续造成比保留系统更大的不可逆损害且未降低锁定或支配，制度提案必须重构。
- 跨物理、生态、网络与社会领域的共同词汇只能形成隐喻，无法说明对象同构与机制迁移条件，跨域主张应删除。

- 若出现跨理论、跨系统可复现且不能由脚本、控制者或观察投射解释的主体证据组合，系统—主体接口必须重开。

11.3 对文明跃迁的反向校准

系统与复杂性科学要求上游哲学把“系统不是主体”改为“系统性与整体功能不足以证明主体”，并禁止用熵、资源有限或动力学不可逆证明意义只能来源于有限性。心灵与生命研究不得把自组织、自维持、记忆、适应、信息整合或韧性单独当作生命、体验或主体连续性的测试；认识论必须把模型完整性、机制成熟度和证据强度分开，证据等级不机械触发制度权限。

政治治理必须把动力学临界与政治紧急状态分开，把高尺度协调需求与中央主权分开，把因果位置与责任归属分开。跨学科情景推演只能构建竞争性因果假设，不能把反馈图当作心智读取；第一人陈述、可观察行为、利益位置和潜在动机必须分栏，未观察状态应允许拒答。高级 AI 系统保证应绑定对象、环境与规模，边界变化、群体行为、共同故障域和权限传播触发重新验证。任何指标化分配不得转为人格、意义、基本权利或政治资格总账。

11.4 开放研究议程

- 建立跨尺度因果图与责任交接图的共同表示，比较其相对普通事故树与组织图的增量。
- 在可控多智能体环境中测试拥塞、串谋、权限扩散、退出者待遇与无代理者排斥。
- 研究预测进入制度后的自我实现、自我消解、数据污染与异议行为，并识别何时反身性可忽略。
- 构建云、身份、支付、能源与通信共同依赖的异构备用和迁移压力测试。
- 发展对信任、专业能力、生态恢复与地方知识等慢变量的长期观察设计。
- 比较集中、多中心与嵌套治理在不同级联速度、知识分布和权利结构中的条件性绩效。
- 把恢复、维持状态域、适应与转型分开测量，识别同一结构提高一类韧性却降低另一类韧性的条件。
- 研究试点扩散、采购标准、数据网络效应和人才迁移如何形成或解除制度锁定。
- 建立跨领域涌现主张复核程序，明确模型同构、机制迁移与外推失败。
- 比较系统退出、分拆、迁移、降级和长期转型的现实成本与不可逆余项。
- 为未来机器群体主体候选建立不以功能绩效替体验证据、也不以人工来源永久排除的联合证据程序。

结论：整体能力越强，推导边界越需清楚

学科重新对账必须同时完成对象辨明与关系深入。各学科可以保持边界，但不能让现实因果在边界处停止；对多维关系、反馈与下一轮目的的研究，决定专业知识能否共同进入文明的整体判断。

AGI 时代要求系统科学重新对账的，不是整体思维是否仍然重要，而是整体能力进入现实之后，谁定义系统、谁被排除、模型如何改变对象、局部优化如何跨尺度回流，以及功能统一为何不能静默继承主体性与公共权力。复杂性不是神秘化的许可证，也不是放弃判断的理由。它要求结论绑定对象、边界、尺度、机制、证据和失效条件，也要求承认一个更困难的事实：局部判断可以分别正确，聚合、耦合与反馈后的整体后果仍可能失灵。

本文由此提出七项收束：涌现必须接受机制—尺度—反事实审计；因果、模型、责任和法律边界分别证明；局部理性与单一总体最优同时受限；自组织不自授权、韧性不自证良善、临界信号不自生紧急权力；反身系统的证据必须版本化；分布因果向责任研究交接而不替其裁决；系统应具备退出、分拆与转型的真实路径。

复杂性科学能够说明整体模式如何形成、何时改变以及模型在哪些边界内失效；它不能单独说明整体是否具有体验、秩序是否正当、风险应由谁承担或权力应归谁。系统科学的边界因此不是墙，也不是现实问题的终点：它保护机制判断的独立性，同时标明证据在哪里止步、责任与授权在哪里必须交接。哲学前提负责公开价值关切和非蕴含边界，专业事实负责约束乃至修订这些前提，政治与法学再承担授权、权利、责任和制度选择。任何一层都不得以“复杂性”“意义”或“体系一致性”取代下一层的独立论证。

锁定结论

系统可以形成整体能力，不能因此取得整体主体地位或公共主权；协调可以跨越尺度，不能因此静默继承授权。

复杂性越高，越需要边界可见、证据可修订、权力可争议、后果可修复、系统可退居。

续读与共同研究入口

本文在系统科学与复杂性科学处只回答：对象与边界如何划定，反馈、时滞、涌现、级联、反身性和共享依赖怎样改变整体后果，以及组件分别合理时为什么仍可能共同失灵。续读不按文件编号排列，而按问题继续；当问题进入技术保证、权利与授权、资源与分配、社会关系、主体证据和现实实验时，其他学科、制度与实践必须重新进入。

上 | 如果继续追问：为什么系统问题最终成为文明问题

续读《AGI 时代的学科重新对账——学科边界、相互作用与共同研究议程》V2.4。它说明为什么专业分工仍是知识精确的前提，却不能成为现实因果链的终点：技术、制度、经济、社会、权力与意义条件共同运行后的结果，才决定一个方向或方案是否整体成立。若进一步追问文明为什么需要跃迁、最终要守住什么和抵达何处，可进入《文明跃迁：AGI 时代的文明之路（全五卷）》与理论体系总入口。

文明跃迁五卷书：<https://www.civitas.top/documents/p02-five-001/>

文明跃迁体系公众阅读与传播总入口：<https://www.civitas.top/documents/p01-entry-006/>

文明跃迁完整文库：<https://www.civitas.top/library/>

下 | 如果继续追问：系统机制怎样变成可验证的工程、推演与实验

可继续进入 P11 人工智能与计算机专业稿，以及 Human-COS、AGI-COS、风险推演、公共窗口和国家级综合实验等现行文稿，检验联合配置、共享依赖、故障注入、熔断、退居、责任交接与慢变量监测怎样在真实系统中被实现。这里的“向下”不是把复杂性理论直接写成治理参数：工程试验、制度原型和现实结果必须拥有独立证据、授权与失败条件，并有权反向修改本文。相关现行版本以文明跃迁文库为准。

横向 | 当系统科学的证据已经止步，谁必须重新进入

- 人工智能与计算机科学：继续回答模型、智能体、部署服务与基础设施究竟怎样构造，系统保证、权限、停止、回滚和退役是否可验证。
- 政治学与伦理／法学：继续回答事实权力、紧急权限、合法性、权利、责任、救济与主权，防止高尺度协调需求被静默换成权源。
- 经济学、政治经济学与价值理论：继续研究局部优化、共享资源、算力与平台集中、外部性、分配和退出成本，检验总体最优怎样改变真实激励。
- 社会学、人类学与文明研究：继续观察指标、分类、平台和 AI 代理怎样改变行为、关系、地方知识、信任与制度实践，防止模型中的“系统”吞并真实生活世界。
- 认识论、生命科学与认知／心灵研究：当整体功能被进一步推向主体、体验、连续性或意义承担判断时，重新回答什么证据足以支持这些命题。

反向 | 哪些竞争路线最值得继续检验

相反性必须保留。强集中路线会追问：高速级联和共同故障是否要求更集中、更统一的感知与控制；严格优化路线会追问：只要目标定义足够好，单一全局模型是否比多模型、边界重开和地方裁量更有效；韧性工程路线会追问：在关键基础设施中，维持与恢复是否应优先于退出、分拆和转型；去中心化路线则会追问：集中本身是否是共享故障与支配的主要来源。本文不预先选定集中或分散、统一或多元。若这些竞争结构在相同后果等级、权利约束与时间尺度下持续给出更好的安全、恢复、分布和可问责结果，本文关于模型多元、退居、嵌套治理或边界重开的主张就必须收缩、重写甚至放弃。

向前 | 如果希望从阅读进入共同研究与现实验证

- 建立跨尺度因果图、共享依赖图与责任交接图的共同表示，比较它们相对普通事故树、组织图和单层网络模型是否真的增加解释与预测。
- 在多智能体与公共资源环境中做可复现故障注入，检验局部忠实、重试、共享凭证、共同云／能源依赖如何产生拥塞、级联、串谋或无代理者排斥，并公开负结果。
- 建立快变量—慢变量联合数据集，长期追踪信任、专业能力、生态恢复、地方知识、供应商依赖和退出能力，验证短期成功何时变成长期锁定。
- 对同一跨域问题并行比较集中、多中心、嵌套与混合治理，预注册何种级联速度、知识分布、外部性与权利条件会支持哪一种结构。

- 对“系统退居”开展现实成本测试：比较停止、降级、分拆、迁移、替代接管和长期转型的服务损失、恢复时间、锁定下降与遗留责任。
- 把指标反身性做成可检验问题：比较指标公开前后、进入资源分配前后及改变激励后的行为与数据生成，识别何时反身性显著、何时可以忽略。
- 如果新的机制、数据或反例证明某些边界与尺度审查长期没有增量，应删除它们；如果新的主体证据出现，则把系统—主体接口交还给相邻学科重新判断。

最有价值的参与，不是把“复杂系统”当作新的总解释，而是用机制模型、跨尺度数据、失败案例、负结果和真实工程验证迫使本文收缩、纠错、重构甚至失败。

续读不是为了把读者留在体系里，而是让不同专业从自己的位置继续向前。不是为了让更多人读完文明跃迁，而是为了让更多人一起完成文明跃迁。

文明未来链 | 如果局部正确仍可能整体失败，什么样的文明能把反馈、退出与长期后果真正活出来？

愿景不是证据。下面的入口不用于证明本文已经成立，而是把本学科判断放进一个可以运行、生活与形成的未来文明中继续检验：读者既可以据此形成方向与向往，也可以拒绝现有图景，并用更强的专业事实、制度方案或现实结果改变它。

继续追问	建议进入	此时能看见什么	必须守住的边界
完整未来图景	《活的文明》三部曲统一出版合编版	把复杂系统的边界、反馈和外部性重新放回普通生活与文明形成，观察整体能力究竟服务了什么。	叙事中的整体协调不能证明系统存在唯一最优。
可退居的系统怎样运行	《活的文明·第一卷：未来文明运行原理》	继续看资源、治理、AI与公共基础设施怎样在可停止、可替代和可退出的条件下形成共同能力。	运行秩序不能替工程测试证明韧性、可逆或安全。
系统怎样从试验长成文明	《活的文明·第三卷：新文明的形成》	观察局部试验、社区、制度与基础设施怎样扩散、锁定或被修正。	扩散成功不等于整体正确，历史形成也不自动生成规范正当性。
复杂文明的系统承载	《星云全景版》	进一步检验整体显影、跨尺度协调、慢变量与多中心结构怎样进入稳定文明。	星云不拥有统一意志或执行主权；全景是待验证系统图景。
机器侧退居与边界	《AGI-COS 全系列总览》	把系统保证、权限、回滚与退役放入机器侧工程约束。	工程安全不能自动证明社会整体良善。

固定回流：[文明跃迁完整文库](#) | [公众阅读与传播总入口](#)

Human-COS 等尚未在统一公共映射中取得无歧义稳定单项页的入口，本稿只写明职责，不猜测 URL；需要时从完整文库进入当前公开版本。

附录 A 来源—主张边界登记

下表不是权威排名或证据总分。每组来源只承担有限结论，并同时登记不能由它单独推出的判断。经典地位、模型优雅、案例规模与引用数量均不能替代对象、假设和外部效度。A／B／C 用于区分经验与前沿证据，F 用于公开哲学／规范承诺；F 不按科学证据等级参与“加权”。来源核验截至 2026 年 8 月 24 日。

附表 A 26 组来源、原创主张与规范承诺的可支持范围

序号	来源／对象	证据地位	可支持的有限结论	单独不足以支持
[1-5]	开放系统、控制、层级与尺度	A：经典理论／方法	开放、反馈、近可分解与尺度差异的条件性工具。	拥有模型即拥有完整现实、目标设定权或主权。
[6]	非平衡自组织	A：物理化学理论	开放受驱非线性系统可形成耗散结构。	社会秩序、自由、合法性或文明意义。
[7]	自组织临界	A：构造模型	特定慢驱动—耗散—阈值模型中的 SOC 与尺度事件。	幂律或 1/f 噪声单独确诊 SOC。
[8]	复杂适应系统	A：研究框架	适应行动者、聚合、非线性、流与多样性。	适应等于进步、正当或主体资格。
[9-10]	小世界与优先连接	A：构造模型	高聚类短路径及增长＋优先连接的生成可能。	现实网络必由该机制产生或具有特定治理结论。
[11]	无标度网络识别争论	B：竞争定义／检验	尾部定义、拟合门槛与替代分布必须公开。	一句话终结现实网络是否无标度。
[12-13]	复杂性—稳定性	A 模型／B 外推	随机模型和作用结构可改变局部稳定条件。	现实生态、金融或文明复杂必稳／必不稳。
[14-17]	韧性、适应、转型与制度稳健性	A：概念／框架	扰动响应需绑定对象、功能、时间窗和制度关系。	高韧性天然良善或转型代价已被分配。
[18]	临界减速综述	A 机制／B 外推	特定分岔附近可出现候选预警统计特征。	普遍、无误报的文明临界报警器。
[19]	整湖预警实验	A：受控生态经验	一个有对照的整湖实验中部分指标先于转变。	所有社会或文明突变都有相同信号。
[20-21]	预警检测边界	A：方法研究	数据、模型与条件选择可产生误报漏报。	预警研究无价值或无需任何预防。
[23-25]	指标与排名反身性	A 窄域／B 外推	高后果指标与公开排名可能改变被测过程。	任何量化、评估或审计必然失败。
[26-28]	弱涌现与因果涌现争论	B：哲学／形式模型	宏观解释或度量在特定模型中可能有增量。	宏观实体普遍本体独立或天然优先治理。

序号	来源／对象	证据地位	可支持的有限结论	单独不足以支持
[29 -30,49]	机器行为、系统 安全与正常事故	A 框架／B 外推	机器社会效应需跨学 科研究；复杂交互、紧耦 合、控制与组织互动都可 能进入系统失效。	机器已有意识； 事故不可避免；安全 等于伦理与合法性。
[31]	隔离动态模型	A：构造模型	局部偏好可产生宏观分离，宏 观不能直接反推动机。	现实隔离只有偏好而无权 力、制度与历史。
[32-34]	公地与多中心治理	A 案例／B 制度外 推	开放获取风险与条件性自组 织、多中心安排。	私有化、中央化或去中心 化的普遍优越。
[35]	路径依赖与技术锁定	A：形式模型／B 现 实识别	递增收益条件下历史事件可改 变技术路径。	任何既有投资都证明现实 已不可逆锁定。
[36]	NK 景观与混沌边缘	A 构造／B 外推	特定随机布尔与适应景观模型 的性能结构。	生命、组织或 AGI 普遍应 在混沌边缘。
[37]	2003 年美加大停电	A：官方事故调查	植被、软件告警、态势感知、 协调和保护动作共同形成级 联。	单一拓扑模型解释全部事 故。
[38]	2021 年德州寒潮	A：官方事故调查	严寒、发电与燃料供应及气— 电依赖造成失效与减载。	与 2003 年属于同一种经典 传输级联。
[39-40]	金融网络传播	A 模型／A 窄域经 验／B 外推	连接在冲击规模变化时可呈稳 健—脆弱反转；同业暴露可传 播。	连接密度单独决定所有金 融危机。
[41]	疫情—行为反馈	A 模型／B 政策外 推	风险信息可改变行为并反馈传 播。	模型已量化现实信任、污 名或强制政策效果。
[42-46]	气候临界与 AMOC 争论	A 候选机制／B 阈 值与时点	多种非线性机制与模型分歧值 得进入预防和模型比较。	确定倒计时、确定崩溃或 确定不崩溃。
本文原创	机制审计、双重反最优、 跨层防火墙	C：方法／规范桥 接	形成可观察问题、推导禁区 and 可修订接口。	复杂性科学已经证明其全 部制度结论。
本文原创	系统退居与复杂性主权批 判	C：制度提案	退出、分拆、迁移、授权和主 权需要独立审查。	唯一治理方案或系统永不 可能成为主体。
本 文 公 开 承 诺	意义多元、主体 开放与跨层非转换	F：哲学／ 规范承诺	说明复杂性研究为何 不得静默生成权利、人 格、授权、主权或意义总 账。	把规范承诺伪装 成复杂性科学事实、 统一价值总分或固定 治理方案。

附录 B 证据、外推与跨层审计清单

附表 B-1 A/B/C/F 论证身份及允许使用

证据地位	构成	允许使用	禁止越界
A 较稳定支持	明确对象、假设和边界内的理论结果；受控实验；多来源事实；官方事故调查。	在声明范围内解释、验证或设计定向监测。	自动外推到文明尺度、规范正当性或政治授权。
B 竞争性解释	模型依赖、定义争议、领域差异、有限外部效度或竞争机制并存。	比较模型、敏感性、情景、领域试验与持续更新。	写成共识、确定阈值、普遍定律或唯一因果。
C 思辨连接	未来 AGI 情景、规范桥接、制度构造或尚无现实对象的主体假说。	概念压力测试、研究设计、最低预防与证据保存。	写成已经发生、已经证实或自动产生权利与主权。
F 哲学／ 规范承诺	公开署名的价值前提、非蕴含边界与公共正当性承诺。	公开前提、最强反例、竞争规范、现实后果与可修订条件。	写成经验事实、复杂性定律、统一价值总分或不可质疑的制度权源。

B.1 外推前的十项检查

- 对象、环境、边界、状态变量、空间尺度与时间窗是否明确？
- 原始来源给出的结果是数学证明、模拟、实验、观察、事故调查还是解释框架？
- 现实对象与原模型在机制、网络、适应性、权力和反身性上是否同构？
- 关键概念是否发生偷换，例如把 SOC、混沌边缘、临界减速和一般突变混为一谈？
- 是否比较共同外因、简单聚合、测量伪象、选择偏差和替代分布？
- 部署或公开模型是否改变行为、资源和下一轮数据生成？
- 总体结果是否同时展示分布、不可退出者、慢变量和不可补偿后果？
- 经验机制是否被静默换成价值判断、法律责任、授权、权利或主权？
- 是否存在可区分的反事实、停止条件、修订触发和退出路径？
- 读者能否在不接受文明跃迁其余部分的情况下独立反驳本文？

B.2 系统—主体接口再审

系统层面的记忆、协调、自我修复、目标保持和整体行动，只能证明相应功能统一；这些性质无论单独还是合取，都不足以证明第一人称体验、本体主体、法律人格、代表关系或公共主权。系统性本身既不证明也不排除主体性。出现新的体验报告、连续性证据或自身后果关系时，应保存状态、比较操控与脚本解释、进行跨系统复现，并由心灵、生命、认识论、伦理、法律与政治程序分别审查。

复杂性研究可以观察哪些系统条件支持或破坏主体的生活实践、关系连续性与意义形成空间，却不能把持久性、目标一致、信息整合、适应成功或总体绩效解释为意义本身。系统能够维持目标，不等于系统拥有意义；系统提高主体福利，也不等于取得意义解释权。

参考文献

- [1] von Bertalanffy L. The theory of open systems in physics and biology. *Science*. 1950;111(2872):23–29. <https://doi.org/10.1126/science.111.2872.23>
- [2] Wiener N. *Cybernetics: Or Control and Communication in the Animal and the Machine*. 2nd ed. Cambridge, MA: MIT Press; 1961. First edition 1948. <https://mitpress.mit.edu/9780262537841/>
- [3] Ashby WR. *An Introduction to Cybernetics*. London: Chapman & Hall; 1956. Digital edition: <https://ashby.info/>
- [4] Simon HA. The architecture of complexity. *Proceedings of the American Philosophical Society*. 1962;106(6):467–482. <https://doi.org/10.2307/985254>
- [5] Anderson PW. More is different. *Science*. 1972;177(4047):393–396. <https://doi.org/10.1126/science.177.4047.393>
- [6] Nicolis G, Prigogine I. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. New York: Wiley; 1977.
- [7] Bak P, Tang C, Wiesenfeld K. Self-organized criticality: an explanation of 1/f noise. *Physical Review Letters*. 1987;59:381–384. <https://doi.org/10.1103/PhysRevLett.59.381>; Bak P, Tang C, Wiesenfeld K. Self-organized criticality. *Physical Review A*. 1988;38:364–374. <https://doi.org/10.1103/PhysRevA.38.364>
- [8] Holland JH. *Complex adaptive systems*. *Daedalus*. 1992;121(1):17–30. <https://www.jstor.org/stable/20025416>
- [9] Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature*. 1998;393:440–442. <https://doi.org/10.1038/30918>
- [10] Barabási A-L, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509–512. <https://doi.org/10.1126/science.286.5439.509>
- [11] Broido AD, Clauset A. Scale-free networks are rare. *Nature Communications*. 2019;10:1017. <https://doi.org/10.1038/s41467-019-08746-5>; Voitalov I, van der Hoorn P, van der Hofstad R, Krioukov D. Scale-free networks well done. *Physical Review Research*. 2019;1:033034. <https://doi.org/10.1103/PhysRevResearch.1.033034>
- [12] May RM. Will a large complex system be stable? *Nature*. 1972;238:413–414. <https://doi.org/10.1038/238413a0>
- [13] Allesina S, Tang S. Stability criteria for complex ecosystems. *Nature*. 2012;483:205–208. <https://doi.org/10.1038/nature10832>
- [14] Holling CS. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*. 1973;4:1–23. <https://doi.org/10.1146/annurev.es.04.110173.000245>
- [15] Walker B, Holling CS, Carpenter SR, Kinzig A. Resilience, adaptability and transformability in social–ecological systems. *Ecology and Society*. 2004;9(2):5. <https://doi.org/10.5751/ES-00650-090205>
- [16] Anderies JM, Janssen MA, Ostrom E. A framework to analyze the robustness of social-ecological systems from an institutional perspective. *Ecology and Society*. 2004;9(1):18. <https://doi.org/10.5751/ES-00610-090118>
- [17] Ostrom E. A diagnostic approach for going beyond panaceas. *Proceedings of the National Academy of Sciences*. 2007;104(39):15181–15187. <https://doi.org/10.1073/pnas.0702288104>
- [18] Scheffer M, Bascompte J, Brock WA, et al. Early-warning signals for critical transitions. *Nature*. 2009;461:53–59. <https://doi.org/10.1038/nature08227>

- [19] Carpenter SR, Cole JJ, Pace ML, et al. Early warnings of regime shifts: a whole-ecosystem experiment. *Science*. 2011;332:1079–1082. <https://doi.org/10.1126/science.1203672>
- [20] Boettiger C, Hastings A. Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface*. 2012;9:2527–2539. <https://doi.org/10.1098/rsif.2012.0125>
- [21] Boettiger C, Hastings A. Early warning signals and the prosecutor’s fallacy. *Proceedings of the Royal Society B*. 2013;280:20131372. <https://doi.org/10.1098/rspb.2013.1372>
- [22] Conant RC, Ashby WR. Every good regulator of a system must be a model of that system. *International Journal of Systems Science*. 1970;1(2):89–97. <https://doi.org/10.1080/00207727008920220>
- [23] Goodhart CAE. Problems of monetary management: the U.K. experience. In: *Papers in Monetary Economics*. Reserve Bank of Australia Conference Volume; 1975/1976. Bibliographic record: <https://www.rba.gov.au/publications/rdp/1990/9013/conference-volumes.html>
- [24] Campbell DT. Assessing the impact of planned social change. 1976; reprinted in *Journal of MultiDisciplinary Evaluation*. 2011;7(15):3–43. <https://doi.org/10.56645/jmde.v7i15.297>
- [25] Espeland WN, Sauder M. Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*. 2007;113(1):1–40. <https://doi.org/10.1086/517897>
- [26] Bedau MA. Weak emergence. *Philosophical Perspectives*. 1997;11:375–399. <https://doi.org/10.1111/0029-4624.31.s11.17>
- [27] Hoel EP, Albantakis L, Tononi G. Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences*. 2013;110(49):19790–19795. <https://doi.org/10.1073/pnas.1314922110>
- [28] Eberhardt F, Lee L-L. Causal emergence: when distortions in a map obscure the territory. *Philosophies*. 2022;7(2):30. <https://doi.org/10.3390/philosophies7020030>
- [29] Rahwan I, Cebrian M, Obradovich N, et al. Machine behaviour. *Nature*. 2019;568:477–486. <https://doi.org/10.1038/s41586-019-1138-y>
- [30] Leveson NG. *Engineering a Safer World: Systems Thinking Applied to Safety*. Cambridge, MA: MIT Press; 2012. <https://doi.org/10.7551/mitpress/8179.001.0001>
- [31] Schelling TC. Dynamic models of segregation. *Journal of Mathematical Sociology*. 1971;1(2):143–186. <https://doi.org/10.1080/0022250X.1971.9989794>
- [32] Hardin G. The tragedy of the commons. *Science*. 1968;162:1243–1248. <https://doi.org/10.1126/science.162.3859.1243>
- [33] Ostrom E. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press; 1990. <https://doi.org/10.1017/CBO9780511807763>
- [34] Ostrom E. Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change*. 2010;20(4):550–557. <https://doi.org/10.1016/j.gloenvcha.2010.07.004>
- [35] Arthur WB. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*. 1989;99(394):116–131. <https://doi.org/10.2307/2234208>
- [36] Kauffman SA. *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press; 1993. <https://doi.org/10.1093/oso/9780195079517.001.0001>
- [37] U.S.–Canada Power System Outage Task Force. *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. April 2004. <https://www.ferc.gov/industries-data/electric/overview/electric-reliability/power-blackout>
- [38] Federal Energy Regulatory Commission, North American Electric Reliability Corporation, and Regional Entities. *The February 2021 Cold Weather Outages in Texas and the South Central United States*. November 2021. <https://www.ferc.gov/media/february-2021-cold-weather-outages-texas-and-south-central-united-states-ferc-nerc-and>
- [39] Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. Systemic risk and stability in financial networks. *American Economic Review*. 2015;105(2):564–608. <https://doi.org/10.1257/aer.20130456>
- [40] Iyer R, Peydró J-L. Interbank contagion at work: evidence from a natural experiment. *Review of Financial Studies*. 2011;24(4):1337–1377. <https://doi.org/10.1093/rfs/hhp105>
- [41] Funk S, Gilad E, Watkins C, Jansen VAA. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*. 2009;106(16):6872–6877. <https://doi.org/10.1073/pnas.0810762106>

- [42] Lenton TM, Held H, Kriegler E, et al. Tipping elements in the Earth' s climate system. *Proceedings of the National Academy of Sciences*. 2008;105(6):1786–1793. <https://doi.org/10.1073/pnas.0705414105>
- [43] Intergovernmental Panel on Climate Change. *Climate Change 2021: The Physical Science Basis*. Working Group I Contribution to the Sixth Assessment Report. Cambridge University Press; 2021. <https://www.ipcc.ch/report/ar6/wg1/>
- [44] Armstrong McKay DI, Staal A, Abrams JF, et al. Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science*. 2022;377:eabn7950. <https://doi.org/10.1126/science.abn7950>
- [45] Ditlevsen P, Ditlevsen S. Warning of a forthcoming collapse of the Atlantic meridional overturning circulation. *Nature Communications*. 2023;14:4254. <https://doi.org/10.1038/s41467-023-39810-w>; Author Correction. 2025. <https://doi.org/10.1038/s41467-025-63201-y>
- [46] Baker JA, et al. Continued Atlantic overturning circulation even under climate extremes. *Nature*. 2025. <https://doi.org/10.1038/s41586-024-08544-0>
- [47] Sterman JD. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science*. 1989;35(3):321–339. <https://doi.org/10.1287/mnsc.35.3.321>
- [48] Bar-Yam Y. Multiscale variety in complex systems. *Complexity*. 2004;9(4):37–45. <https://doi.org/10.1002/cplx.20014>
- [49] Perrow C. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books; 1984.

生成式人工智能使用声明：本稿在资料与文献线索整理、结构设计、部分文字起草与润色、跨稿一致性检索、引用线索核查和版式检查中使用了生成式人工智能工具。生成式人工智能输出未被作为经验数据、权威来源或独立证据。研究问题、理论命题、规范判断、材料取舍、引文采用和最终定稿均由作者审定；作者对事实准确性、引文真实性、论证及全部结论承担责任。人工智能不列为作者，也不承担学术责任。正式投稿时，将按目标期刊要求补充工具名称、版本、使用日期和具体范围。