

AGI 时代的生态学、地球系统与资源环境重 新对账

——算力扩张、能源资源与文明承载边界

Reckoning with Ecology, Earth Systems, and Resource-Environmental Constraints in
the Age of AGI

Compute Expansion, Energy and Resources, and the Material Boundaries of Civilization

生态／地球系统／资源环境独立外发专业深稿 V1.1

作者：子君赋 | 出品：文明跃迁研究组

证据截点：2026 年 8 月 26 日

摘要

人工智能常被描述为一种可以近乎无限复制的数字能力，但任何现实中的 AI 都依赖芯片、服务器、数据中心、电网、冷却、水、土地、材料、通信网络、建设工程以及设备更新与废弃。AGI 时代真正需要重新对账的，不是“AI 是否耗电”这一单一问题，而是当智能能力逐步成为科研、生产、教育、医疗、治理与安全的共同基础设施以后，计算需求怎样与真实地球系统发生长期耦合：单位计算效率的提高能否抵消总需求扩张；全球占比看似有限的负荷是否在局部电网和流域形成高压；降低一种资源消耗是否会增加另一种环境代价；设备更新、材料与关键矿产怎样进入生命周期影响；以及 AI 自身能否通过电网调度、材料发现、气候与能源优化反过来降低环境压力。

本文从生态学、地球系统科学、能源与电力系统、水资源、工业生态学、生命周期评价、材料流与资源环境研究的成熟专业遗产出发，建立严格的论证分层。A 类包括可核验的数据中心用电、区域负荷、用水、正式统计、生命周期方法与现实基础设施事实；B 类包括 AI 未来能耗、水耗、碳排、反弹效应、硬件寿命和减排潜力等尚依赖情景与假设的竞争判断；C 类包括“AI 能力扩张→算力与基础设施扩张→能源／水／材料压力→地区生态与资源后果→价格、就业与治理反应→反过来改变 AI 扩张”的跨域因果链；F 类则公开承担“任何关于未来智能丰裕的文明想象，都必须接受真实物质世界检验”的规范承诺。本文不把 A 类资源事实自动写成反增长结论，也不让生态学替经济学、政治学、法学或文明方向决定最终价值排序。

截至 2026 年，国际能源署更新判断认为全球数据中心用电由 2025 年的约 485 TWh 可能增长至 2030 年的约 950 TWh，约占届时全球用电 3%，其中 AI 专用数据中心用电增长更快；美国劳伦斯伯克利国家实验室的底层设备模型则给出美国 2030 年数据中心用电参考情景 649 TWh、约占美国总用电 11.8%，并明确给出较宽的不确定区间。[1-3] 这些数字不是文明未来的确定预测，而是说明一个方法论事实：AI 环境影响必须绑定时间、地点、系统边界、能源结构、设备与利用率。水资源研究显示，同一计算工作负载的耗水可因服务器效率、电网水强度、利用率、冷却技术和气候等因素产生超过四个数量级的差异；近期综合研究还提示，减少用水与降低碳排在一些条件下可能发生权衡。[6,10] 因而，单一“每次提示耗多少水”或“每个模型耗多少电”的数字不能承担完整生态判断。

本文最终提出：智能丰裕不是“计算的边际成本下降”即可成立，而必须继续经过电力可得、局部电网承载、水与生态可持续、材料与供应链、生命周期和可恢复性等现实检验；同时，环境约束也不能被误写为文明必须停止技术进步。真正的问题是，能否把 AI 发展从“更大规模”转向“更高真实效用／单位环境代价、更强区域适配、更长硬件生命周期、更高资源循环、更可调度负荷以及对能源和生态系统本身的净改善”。这是一项可反驳、可修订的专业研究议程，而不是以生态之名预先结算文明方向。

关键词：AGI 时代；生态学；地球系统；数据中心；能源系统；水资源；关键矿产；生命周期评价；反弹效应；文明承载

Abstract

Artificial intelligence is often described as a digital capability that can be replicated at very low marginal cost. Yet every real deployment depends on chips, servers, data centres, electricity grids, cooling systems, water, land, materials, networks, construction, and equipment turnover. The relevant environmental question in the age of AGI is therefore not simply whether AI consumes electricity. It is how expanding computation becomes coupled to real Earth systems once advanced AI functions as shared infrastructure across science, production, education, health care, government, and security.

This paper draws on ecology, Earth-system science, energy and power systems, hydrology, industrial ecology, life-cycle assessment, material-flow analysis, and resource-environment research. It distinguishes established or verifiable facts and formal data (A), unsettled mechanisms and scenario-dependent estimates (B), cross-domain structural hypotheses (C), and the author's explicit normative commitments (F). It argues

that unit efficiency cannot be equated with lower aggregate resource use; a small global electricity share can coexist with severe local grid or watershed constraints; reducing water, carbon, land, or material impacts can involve trade-offs; and operational energy alone is insufficient for assessing the life-cycle footprint of AI.

The paper treats “intelligence abundance” as a physical-ecological proposition as well as an economic or technical one. Any claim that advanced AI can create abundance must remain answerable to electricity availability, grid integration, water and ecological conditions, material supply, embodied impacts, equipment turnover, and recoverability. At the same time, environmental constraints do not themselves determine the final social or political choice. AI may also improve power-system flexibility, energy innovation, material discovery, climate analysis, and resource efficiency. The objective is therefore not to oppose technological expansion, but to identify the conditions under which intelligence can scale without externalizing irreversible material and ecological costs.

Keywords: AGI era; ecology; Earth system; data centres; energy systems; water resources; critical minerals; life-cycle assessment; rebound effect; civilizational carrying capacity

研究定位、对象边界与论证纪律

本文是一篇生态／地球系统／资源环境专业深稿。它独立承担的是：AI 与数据中心进入真实物质世界以后，能源、电网、水、土地、材料、关键矿产、制造、排放、废弃物、区域生态和地球系统反馈怎样被测量、解释和约束。它不把“算力”当作纯粹抽象量，也不把环境问题缩成一张碳排清单。本文的最小对象是一条物质链：计算任务由什么硬件完成，硬件在哪里运行，使用什么电力和冷却，电与水从哪里来，设备怎样制造与更新，基础设施占用什么土地和材料，废弃后如何处置，以及这些局部压力进入什么流域、电网、生态系统和全球环境过程。

本文不替人工智能与计算机科学判断芯片、模型、软件和系统性能；不替经济学决定价格、投资、所有权和资源分配；不替政治学与法学决定电网、水、土地和矿产由谁治理；不替伦理学决定代际责任和环境正义的最终规范；也不替文明方向稿决定“发展到哪里”。生态学与资源环境科学的专业责任，是把真实物质条件证明到能够被其他学科重新进入的位置。

表 1 A/B/C/F 论证身份与生态专业边界

身份	本文中的典型内容	能够支持什么	不能自动推出什么
A 较稳定事实/正式材料	数据中心实测／统计用电，电网结构，用水与 WUE，正式资源统计，成熟 LCA 方法	支持限定地点、时间、技术和系统边界下的事实判断	不能自动推出 AI 整体净环境效益或某种政策必然正确
B 前沿/竞争解释	未来 AI 能耗与水耗、设备寿命、反弹效应、净减排潜力、材料替代	比较情景、机制和外部效应	不能写成已经收敛的全球结论
C 跨域结构判断	算力扩张→资源压力→价格/治理→反过来改变 AI 路径；AI 优化→资源效率→环境压力下降	形成可反驳的跨域因果假说	不能伪装成单一学科已直接证明的完整闭环
F 规范承诺	任何“智能丰裕”都须接受物质世界检验；不可逆生态损害不能因数字收益而隐形	说明本文为何要求完整边界和真实承载	不能替代经济、法律、政治或文明最终价值排序

“AGI 时代”不等待一个完成态 AGI

本文使用“AGI 时代”描述一种已经前置发生的基础设施条件：高级 AI 正在把生成、分析、代码、推理、工具调用和部分代理能力嵌入越来越多现实任务，由此推动专用加速计算、数据中心、电网接入、冷却与供应链的扩张。真正 AGI 何时出现、是否形成机器主体性或超级智能，都不是本文成立的前提。若未来 AI 能力、算法效率、硬件架构或需求模式发生重大变化，本文的数量判断必须更新，但“计算必须通过真实物质世界发生”的研究责任不会因此消失。

目录

导论 数字能力越强，越要重新看见它依赖的物质世界
一 成熟专业遗产：生态、地球系统与资源环境已经知道什么
二 研究对象：从一次计算到完整物质链
三 电力：全球占比有限，局部承载可以很高
四 电网：年度电量不是接得上、稳得住与供得起
五 水：一个“AI 耗水数字”为何远远不够
六 碳与气候：电力来源、时间地点与生命周期边界
七 材料与关键矿产：智能基础设施也有上游供应链
八 芯片、制造与隐含环境影响：运行之外的环境账
九 电子废弃物与循环：更快更新是不是唯一效率路径
十 土地、生态与地区承载：全球平均值遮不住地方真实
十一 效率与反弹：单位更省为什么总量仍可能增长
十二 AI 也可能降低环境压力：从负荷到环境工具
十三 多目标权衡：能源、水、碳、材料不能互相冒充
十四 地球系统边界怎样进入 AI 研究，而不变成生态总分
十五 “智能丰裕”的现实条件：从数字可能到文明红利
十六 跨学科交接：事实证明以后谁必须重新进入
十七 什么事实能够改变本文：最强反驳与可反驳研究议程
结语 没有脱离地球的智能，只有更好地生活于地球的智能
续读与共同研究入口
附录 A 来源—主张边界登记
附录 B AI 资源环境压力测试矩阵
生成式人工智能使用声明
参考文献
版本说明

导论 | 数字能力越强，越要重新看见它依赖的物质世界

AI 最容易制造的一种错觉，是把“数字”误认为“非物质”。模型权重可以复制，软件可以远程更新，推理接口可以在数秒内扩展到世界各地，于是人们很容易把智能增长想象成只受算法与资本约束。但现实中的每一次训练和推理都发生在服务器上，服务器依赖芯片、机架、供电、冷却、通信和建筑，电力来自具体电网，水来自具体流域，硬件来自具体矿山、冶炼、晶圆厂与供应链，淘汰设备又回到真实废物系统。所谓“云”，从来不是没有地点。

国际能源署 2025 年《Energy and AI》估计，全球数据中心 2024 年约消耗 415 TWh 电力，约占全球用电 1.5%；其 2026 年更新把 2025 年用电估计为约 485 TWh，并维持到 2030 年约 950 TWh 的中央情景，约占全球用电 3%，同时指出 AI 专用数据中心用电增长速度明显更快。[1-2] 这些数字既不支持“AI 将吃掉全球电力”的夸张，也不支持“3%所以无关紧要”的轻率。电力系统具有强烈空间性：同样的全球占比可以集中于少数节点、城市和区域，形成完全不同的建设周期、价格、排放与可靠性压力。

这种“总量不等于局部”的结构已经有现实证据。爱尔兰中央统计局 2026 年发布的 2025 年度数据表明，数据中心已占全国计量电力消费 23%，而 2015 年为 5%。[36] 美国弗吉尼亚州 2024 年 JLARC 研究则指出，数据中心是未来用电需求快速增长的主要驱动，满足不受约束需求将要求极大规模新增发电和输电设施。[37] 两个案例都不能外推为全球结论，却足以说明：环境与资源研究必须同时看全球总量和地区承载。

数字能力可以近乎瞬时复制，电网、流域、材料和生态系统不能。

本文因此不从“AI 绿色不绿色”这个总问题开始，而从一系列可以被专业证明的局部问题开始：电有多少、在何时何地；水从哪里取、怎样消耗；材料在什么环节进入；设备多久更新；排放发生在哪里；土地与生态影响怎样分布；效率提高后总需求是否继续增长；AI 用于能源和环境本身能产生多大真实收益。只有这些问题分别被证明，文明才有资格继续讨论所谓“智能丰裕”的真实可行性。

一 | 成熟专业遗产：生态、地球系统与资源环境已经知道什么

1.1 生态学首先研究关系与承载，而不是给文明打总分

生态学关注生物与环境、种群与群落、能量与物质循环、扰动与恢复等关系。它最重要的启示之一，是影响往往取决于地点、尺度、时序和系统状态：相同的资源占用或排放，在不同生态系统中可能产生不同后果。生态研究可以识别承载压力、脆弱性和不可逆风险，却不能仅凭生态指标决定所有社会价值。本文继承这种专业克制：环境事实必须进入文明判断，但不能独占文明判断。

1.2 地球系统科学处理跨尺度耦合与全球稳定条件

Rockström 等提出的行星边界框架、Steffen 等的更新和 Richardson 等 2023 年的进一步评估，都试图把气候、生物圈、土地、淡水、生物地球化学循环和“新实体”等过程放在地球系统稳定性中共同观察。[22-24] 2023 年更新判断六项边界已被跨越；这是一项具有巨大启发力、同时仍持续演进的科学框架，而不是给单个数据中心发放建设许可的阈值表。本文使用它来提醒研究者：局部资源压力可能通过多个地球系统过程累积，但绝不把“行星边界”机械换算成 AI 项目分数。

1.3 气候科学提供的是因果与风险边界，不是单一技术禁令

IPCC 第六次评估综合报告明确，人类活动已经明确导致全球变暖，气候、生态系统、生物多样性与人类系统相互依赖，部分风险随升温增加并可能出现长期或不可逆后果。[25] 这使数据中心电力来源、建设速度和生命周期排放具有真实气候意义，但“有排放”并不能单独决定某项 AI 用途是否应被禁止；需要比较替代路径、系统收益、排放强度和其它社会目标。

1.4 工业生态学与生命周期评价：不要只看运行阶段

生命周期评价（LCA）长期以目标与范围界定、清单分析、影响评价和解释来处理产品或系统从资源开采到生产、使用和末端处置的环境影响。[19–20] 工业生态学则把生产系统放回材料、能源、废物流与产业间循环中理解，强调不能把资源开采、制造和废弃阶段排除在系统之外；金属生命周期研究进一步说明，不同金属的资源与环境负荷具有显著差异，不能只用“总材料质量”代替材料类型与供应链影响。[30–31] ISO 14046 进一步为水足迹提供生命周期原则与要求。[21] OECD 关于“AI footprint”的政策研究也明确区分 AI 计算与设备的直接环境影响，以及 AI 应用产生的间接成本和收益，并建议超越运行能耗、改进测量标准与透明度。[33] 这些成熟方法意味着：只报训练电量，或只报某一次推理的能耗，都不构成完整 AI 环境评价。

1.5 资源环境研究提醒：稀缺会迁移

UNEP 国际资源小组 2024 年《Global Resources Outlook》指出，全球材料使用在过去约 50 年已增加三倍以上，资源开采与加工本身是温室气体、颗粒物健康影响、土地利用、生物多样性损失和水压力的重要来源。[15] AI 并不是全球资源增长的唯一或主要来源，但当智能基础设施扩展时，它会进入本已存在的能源转型、城市建设、矿产和制造竞争。因此，“数字经济”不能被视为摆脱材料经济，而是材料需求发生结构转换。

二 | 研究对象：从一次计算到完整物质链

AI 环境讨论经常把不同对象混在一起：“一个模型”“一次训练”“一次提示”“一家数据中心”“整个数据中心行业”“全部数字基础设施”并不是同一个研究对象。Masanet 等在 2020 年校准全球数据中心能耗估计时就提醒，服务器效率、虚拟化、基础设施效率和工作负载变化会显著改变总量估计。[5] LBNL 2024 年美国数据中心能耗报告又把服务器、存储、网络、设施效率和不同负荷类型放进同一底层核算框架，为后续 2026 更新提供了直接基线。[4] 进入生成式 AI 时代后，这种对象分层更重要。

表 2 AI 环境研究的六层对象与证据停止点

对象层	包括什么	主要证据	不能自动推出
任务/工作负载	一次训练、推理、检索、生成或代理链	任务时间、token/样本、算力、硬件利用率	不能直接代表数据中心年度环境影响
硬件/服务器	GPU/加速器、CPU、存储、网络和服务器配置	功率、性能、寿命、利用率	不能直接代表电网、用水或材料全生命周期
数据中心设施	供配电、冷却、建筑、备用电源与 IT 设备整体	PUE、WUE、负荷曲线、设施用水	不能自动代表地区电网与流域影响
电网/流域/地区	发电组合、输配电、取水、生态和地方基础设施	边际/平均电力、水压力、容量、排放和生态状态	不能直接生成经济或政治正当性
供应链/生命周期	矿产、冶炼、芯片制造、建筑、运输、更新、废弃	材料流、隐含环境影响、废物和循环	不能用运行效率代替
全球地球系统	气候、生物圈、淡水、土地等大尺度过程	累积压力、边界与长期风险	不能机械换算为单项目许可

2.1 四类系统边界必须公开

第一是时间边界：某个小时的峰值、年度用电、设备寿命与 2030 情景完全不同。第二是空间边界：机房、园区、电网节点、流域、国家和全球不能混写。第三是生命周期边界：只看运行电力，还是包括芯片制造、建

筑、运输与废弃。第四是归因边界：数据中心同时承担传统云服务、存储、视频、企业计算和 AI，运营商通常不会公开到足够细的 AI／非 AI 负荷拆分。任何不说明这四个边界的“AI 环境数字”，其可解释性都有限。

2.2 效率指标有用，但不能越权

ISO/IEC 30134-2:2026 将 PUE 定义为数据中心能源使用效率的标准化关键指标，ISO/IEC 30134-9:2022 定义 WUE 用于量化数据中心使用阶段的耗水强度。[34–35] 两者都很重要，却分别只回答设施能源效率与运行阶段水强度。较低 PUE 不等于更低碳；较低 WUE 不等于更低全生命周期水足迹；二者更不能说明一个 AI 服务是否产生足够社会价值。指标必须回到它本来解决的问题。

三 | 电力：全球占比有限，局部承载可以很高

3.1 当前最可靠的数字仍然是区间与情景

IEA 2025 年估计全球数据中心 2024 年用电约 415 TWh，并在基准情景中预测到 2030 年约 945 TWh；2026 年更新把 2025 年估计提升至约 485 TWh、2030 年约 950 TWh，认为 AI 专用数据中心 2025—2030 年用电可能增长约三倍。[1–2] 这类数字应被解释为在明确假设下的能源系统情景，而不是“AGI 必然耗电多少”的事实。硬件效率、AI 采用、资本市场、芯片与变压器供应、建设许可和电网接入都会改变实际路径。

美国 LBNL 2026 年更新报告采用设备出货、单设备年用电、冷却模拟、设施类型与地点等底层数据建模，给出 2030 年 649 TWh 参考情景、约占美国用电 11.8%，并给出 9.5%—15.3% 的情景范围及 521—843 TWh 的更宽综合不确定区间。[3] 这一研究的价值不只在数字更高或更低，而在于它把“预测”拆成可检验假设：安装多少设备、AI 芯片寿命多久、闲置功率与利用率怎样、设施效率如何。

3.2 全球平均值会掩盖局部高浓度

数据中心用电具有明显地理集中。IEA 指出，美国约一半数据中心容量集中在五个区域集群，局部电力市场中的占比远高于全球平均。[1] 爱尔兰 2025 年数据中心占全国计量电力消费 23% 的统计进一步表明，研究者不能用“全球只有几个百分点”否定本地问题。[36] 反过来，也不能拿爱尔兰的 23% 去说明所有国家都会达到相同水平。

3.3 “电够不够”不是一个全球命题

电力系统的真实问题至少包括年度电量、峰值功率、爬坡速率、节点容量、输电与配电、发电结构、可靠性、备用与建设周期。某个国家全年发电量总体充裕，不代表某个园区可以在两年内接入数百兆瓦持续负荷；某个地区可建设大量可再生能源，也不代表发电时段和 AI 负荷时段天然一致。AI 的能源约束因此是“时空可达性”问题，而不是一个全球电量除法。

四 | 电网：年度电量不是接得上、稳得住与供得起

数据中心的建设速度通常快于电网扩建。IEA 指出，数据中心可以在两三年内投运，但发电、输电、变压器和电网规划往往需要更长周期；2026 年更新又强调，变压器、燃气轮机、芯片和 IT 设备供应链瓶颈已经降低近期高增长情景实现概率。[1–2] 因此，“未来有足够发电技术”不能自动证明“某个地区及时接得上”。

4.1 弗吉尼亚案例：负荷增长会重写基础设施路径

Virginia JLARC 2024 研究认为，在不受约束的需求情景下，该州电力需求可能在未来十年翻倍，数据中心是主要驱动；满足需求需要大规模新增发电与输电，并且即使只满足一半不受约束增长也具有明显建设难度。[37] 这是地区规划研究，不是 AI 全球结论。它证明的是：当大型负荷集中出现时，新增数据中心会改变电力系统投资路径和资源组合，环境分析必须把这种间接基础设施结果纳入。

4.2 数据中心也可能成为可调负荷

把数据中心视为永远不可调整的“刚性负荷”同样可能低估技术空间。Granderson 等 2026 年总结了四类可调机制：计算负荷在时间或地点迁移、设施基础设施调整、储能和现场发电，并指出需求灵活性不能替代长期新增发电，却可以缓解近期并网和峰值压力。[12] Colangelo 等在美国亚利桑那州凤凰城一处超大规模云设施中的 256-GPU 集群开展现场演示，通过软件协调在 3 小时峰值期间降低 25% 功率，同时维持其定义的 AI 服务质量保证，且不依赖硬件改造或储能。[11] 这是一个重要正向 A/B 证据：部分 AI 负荷确有电网交互潜力；它不能外推为所有工作负载、所有地区都能如此调节。

表 3 从“耗电量”进入真实电网的五个问题

电力问题	真正问什么	不能被什么替代
年度电量	一年能生产多少电	可能掩盖峰值、节点和季节约束
峰值功率	最紧张时刻需要多少即时容量	不能说明全年能源与排放
节点/输电容量	负荷能否在具体地点接入	不能由全国发电总量替代
灵活性	负荷能否移时、降功率或迁移	依赖任务时延、数据位置、SLA 与安全
可靠性/备用	故障和极端天气下能否持续运行	可能增加现场燃料、储能与额外基础设施

五 | 水：一个“AI 耗水数字”为何远远不够

水问题尤其容易被传播数字误导。数据中心用水至少有三条路径：设施直接冷却消耗；发电过程的间接耗水；芯片、设备和建筑等上游生命周期水足迹。ISO/IEC WUE 只针对数据中心使用阶段的水消耗指标，并不等于全生命周期水足迹。[21,35]

5.1 同一工作负载的耗水可以相差超过四个数量级

Lei、Lu、Shehabi 与 Masanet 对工作负载层耗水的综述与评估发现，不同条件下同一工作负载的水使用差异可超过 10,000 倍，其中每 kWh 服务器用电相关水消耗因素可相差 1000 倍以上，服务器工作负载效率本身也可相差约 10 倍；主要决定因素包括服务器效率、电网水耗因子、利用率、冷却系统、基础设施效率、气候、闲置服务器比例和更新周期。[6] 这项结果直接否定了——一个传播习惯：把某个模型、某个地区、某个冷却系统测得的“每次提示耗水”写成 AI 普遍常数。

Mytton 与 Siddik 等此前的研究已经表明，数据中心水影响需要同时处理直接现场消耗和电力相关间接耗水，并具有明显地域差异。[7-8] 2025 年 Xiao 等对美国 AI 服务器 2024—2030 情景建模进一步给出较宽的水与

碳区间，并明确指出效率、网格脱碳与空间分布是深度不确定性来源。[9] 因此，对于未来 AI 水足迹，模型研究最诚实的表达往往是条件区间而不是单点预言。

5.2 节水与减碳有时会相互冲突

冷却系统并不存在一个脱离地点的“唯一绿色方案”。Chien 等 2026 年综述指出，在一些情况下，降低数据中心用水的策略会增加 CO₂ 当量排放，说明水与碳可能存在负向耦合，需要结合能源结构、气候与设施方案共同优化。[10] 这不是说“不能节水”，而是说明单一 WUE 最低并不自动代表总体环境最好。

5.3 水必须回到流域与时段

同样的 100 万立方米耗水，发生在高水压力流域、丰水区、旱季或雨季，其生态与社会后果不同。水资源研究必须继续进入取水来源、消耗性使用、回用比例、河流与地下水状态、生态需水、季节波动和其它用户竞争。全国平均水量不能替代地方流域承载，企业年度“补水”承诺也不能自动证明取水在时间和空间上已经等价恢复。

六 | 碳与气候：电力来源、时间地点与生命周期边界

6.1 用电量不等于碳排放量

同样 1 kWh 电力的排放取决于电源结构、时间、地点和采用平均还是边际排放因子。IEA 2025 年估计数据中心相关间接 CO₂ 排放约 180 Mt，并在基准情景中随用电增长上升，但同时强调其全球排放占比仍有限，且不同地区电力结构差异显著。[1] 因而，训练或推理能耗必须与具体电网排放因子绑定；把“耗电”直接写成“污染同等增加”是不严谨的。

6.2 运行排放不是完整生命周期排放

芯片制造、服务器、机房建设、变压器、备用系统和频繁硬件更新都具有隐含环境影响（embodied impacts）。Chien 等综述指出，大型 AI 数据中心的硬件与设施制造可能在特定生命周期边界下占据重要甚至过半的排放份额，并讨论通过延长硬件使用、重复利用和低碳设计降低这类影响。[10] Plociennik 等则指出，现阶段 AI 系统系统化 LCA 仍处初期，不能只以运行电力或 CO₂ 当量概括，还应纳入关键原材料、水、资源耗竭等指标。[18] 这些研究的共同方向是扩大系统边界，而不是宣称已有一个统一 AI 碳足迹常数。

6.3 AI 扩大需求，也可能帮助降低其它部门排放

IEA 同时评估了 AI 在电网、建筑、工业、运输、甲烷监测和能源创新中的潜在减排作用，并给出在广泛采用情景下可能出现的大规模减排潜力，但明确指出当前采用障碍、数据、技能、基础设施和反弹效应使这些收益不能自动兑现。[1] 因而，净气候影响必须比较“AI 自身新增排放”与“AI 促成的真实减排”，且避免把理论技术潜力当作已经实现的净收益。

七 | 材料与关键矿产：智能基础设施也有上游供应链

AI 基础设施的物质性不只体现在电和水。服务器、芯片、供配电、储能和网络需要铜、铝、硅及多种关键或战略矿物。IEA 2025 年《Global Critical Minerals Outlook》显示，能源转型本身已推动关键矿物需求快速增长

长，供应集中、项目周期和出口限制都构成现实风险。[13] 2026 年更新及同期关键矿产数据集进一步显示，多种矿物的精炼集中度仍很高，供应增长和加工能力高度集中于少数国家，具体矿种和年份应回到数据集逐项核对。[14,32] AI 因此进入的不是一张空白资源表，而是一个已经被能源转型、电气化和产业安全共同拉动的材料系统。

7.1 数据中心可能在少数战略材料上产生非线性压力

IEA 2025 年《Energy and AI》估计，到 2030 年数据中心对镓的需求可能相当于当时全球供应的约 10%，并指出中国当时约占全球镓精炼的 95%；IEA 2026 年《Global Critical Minerals Outlook》进一步显示，中国作为最大精炼国仍占全球镓供应的 90% 以上。[1,14] 这些数字不意味着“AI 将耗尽镓”。它们说明：即使某种材料在数据中心总质量中占比很小，只要供应链高度集中、扩产受副产品属性或专业加工能力约束、替代路径有限，也可能成为扩张与安全上的瓶颈。总质量不是战略重要性的充分代理指标。

7.2 材料问题必须与替代、回收和技术路线共同分析

材料需求不是固定系数。芯片工艺、封装、光互连、功率电子、冷却、储能和电网设计都会改变材料组合；回收率、二次材料、设备寿命和模块化设计又决定新开采需求。资源研究必须把“需要多少”拆成技术路线与情景，而不能把今天的一张物料清单永久化。相反，任何“新材料会自动解决瓶颈”的说法也必须给出性能、规模、回收、毒性和供应链证据。

八 | 芯片、制造与隐含环境影响：运行之外的环境账

AI 环境研究若只看服务器插座功率，会遗漏半导体制造和设备更新。芯片制造涉及高纯水、能源、化学品、洁净室、上游材料与复杂全球供应链；不同工艺节点、厂区、电力和水条件差异很大。本文不在缺乏同口径数据时给出一个“每块 AI 芯片固定水耗／碳排”，而强调必须把制造阶段纳入 LCA 系统边界。

Plociennik 等 2025 年明确指出，AI 生命周期评价仍缺少成熟统一方法，现有能耗计算器通常覆盖运行能源和碳换算，却不足以处理关键原材料、硬件生产、维护与终端设备等完整影响。[18] OECD 2022 也建议建立更透明、标准化和可比较的 AI 环境测量体系，并区分直接与间接环境作用。[33] 这两项工作共同说明，今天最重要的不是追求一个看似精确的“AI 总环境分”，而是把缺失数据和边界公开。

8.1 硬件更新周期本身就是环境变量

更快的芯片可以降低单位任务电耗，但如果经济最优使服务器更新周期持续缩短，上游制造和废弃物负担可能同步增加；反过来，过度延长低效率设备寿命也可能增加运行期电耗。最优更新周期因此是生命周期问题，不是“越新越绿”或“越旧越环保”的口号。Chien 等讨论了重复利用或较低性能硬件在部分推理场景中降低隐含排放（embodied emissions）的可能性，但适用范围仍依任务性能与电力条件而定。[10]

九 | 电子废弃物与循环：更快更新是不是唯一效率路径

全球电子废弃物本已是快速增长的废物流。ITU/UNITAR《Global E-waste Monitor 2024》估计，2022 年全球产生约 620 亿千克电子废弃物，仅 22.3% 被记录为正式收集和回收，并预计在既有趋势下 2030 年总量继续显著增长。[16] AI 服务器只是其中一个新增部分，不能把全球 e-waste 全部归因于 AI。

Wang 等 2024 年使用材料流与计算能力情景估计，生成式 AI 相关电子废弃物在 2020—2030 年不同情景下累计可能达到 120 万—500 万吨，并认为循环经济策略有可能显著降低这一废物流。[17] 这是 B 类情景研究，不应写成“2030 必然产生 500 万吨 AI 废物”。它真正提出的研究问题是：服务器更新速度、地缘供应限制、设备再利用、模块维修、回收和跨代兼容怎样共同决定物质结果。

9.1 循环不是把废物末端回收就结束

更成熟的循环路径包括：延长可用设备寿命、将高端训练硬件退居到推理或低负载任务、模块化升级、维修、部件再使用、提高材料回收率以及设计阶段降低难回收复合材料。循环策略必须与网络安全、数据销毁、可靠性和性能要求同时满足；生态学不能因为“循环更好”就忽略安全和服务连续性，工程也不能因为“淘汰更快更高效”就忽略材料后果。

十 | 土地、生态与地区承载：全球平均值遮不住地方真实

数据中心园区本身可能只占有限土地，但它所带动的变电站、输电、发电、储能、道路和供水设施会扩大空间足迹。选址还可能进入湿地、农地、野生生境、历史景观或人口密集区。Virginia JLARC 2024 研究把土地、自然资源、居民噪声、地方税收和能源基础设施放在同一个地区分析中，正说明“数据中心环境影响”不应只剩电表。[37]

10.1 生态影响必须从“地点”开始

生态系统并不按企业组织边界运行。一个设施的现场排放很低，若新增输电穿越敏感栖息地、取水改变枯水期流量、备用发电增加地方空气污染，真实后果仍然存在。IPBES 2019 年全球评估指出，自 1970 年以来，土地利用变化对陆地与淡水生态系统自然状态造成了最大的相对负面直接影响。[26] 这一全球背景说明基础设施选址、配套输电与间接土地改变为什么必须进入生态审查，但它不能反向证明某个具体数据中心已经造成生物多样性损失；项目归因仍须依靠地点、基线、影响路径和累积效应证据。反过来，在棕地、既有工业区、低水压力地区并采用低影响能源和循环水，环境压力可能明显降低。生态研究最重要的不是宣布数据中心“有害/无害”，而是把环境后果定位到具体空间过程。

10.2 紧凑地区会更早暴露承载边界

新加坡 2024 年《Green Data Centre Roadmap》明确承认数据中心是电力和水资源密集型基础设施，并在国土与资源条件约束下提出近期开启至少 300 MW 新增容量、以效率和绿色能源部署作为扩展条件。[38] 这不是一个可复制到所有国家的政策模板，但它说明地区资源条件会直接塑造 AI 基础设施规模、技术与治理路径。

十一 | 效率与反弹：单位更省为什么总量仍可能增长

AI 环境讨论中最容易混淆两个方向：单位任务资源强度和总资源使用。硬件、模型、量化、稀疏化、缓存和设施效率可以让同一任务更省电；但效率降低成本后，使用人群、任务频率、上下文长度、视频生成、代理链和新应用可能同步增长。单位更省与总量下降之间没有自动等号。

11.1 反弹效应是竞争性经验问题，不是定律

能源经济学对反弹效应已有长期争论。Sorrell 2009 年综述指出，所谓“Jevons 悖论”的强版本难以经验确认，但效率提升引发部分反弹具有理论与实证基础，其规模取决于价格、收入、技术和市场结构。[27] 因此，本文不把“AI 更高效→必然用更多电”写成定律，也不把“每 token 能耗下降→行业总耗能一定下降”写成定律。

IEA 2026 年更新给出了一个现实并存状态：AI 任务单位功耗正在快速下降，但 AI 使用和高能耗应用扩张更快，所以数据中心总用电仍被预计继续增长。[2] 这正是为什么环境研究必须同时报告单位效率、利用率、任务规模、总服务量和总资源使用。

表 4 效率改进为什么不能直接换算总环境结果

效率变化	直接收益	需要实证检验的反弹／权衡
单位硬件效率提高	相同工作负载耗电下降	设备更便宜/性能更强后部署量上升
模型推理效率提高	同一输出所需算力下降	更长上下文、更频繁调用、代理多步执行
PUE/WUE 改善	设施每单位 IT 负载资源下降	园区总容量扩张抵消单位改善
低碳电力增加	单位电力排放下降	总负荷增长、其它用电挤出、材料建设增加
设备寿命延长	减少制造与废弃	旧硬件运行效率低导致使用期电耗增加

十二 | AI 也可能降低环境压力：从负荷到环境工具

如果一篇环境稿只列 AI 的资源成本，它同样会失真。可持续 AI 研究已经明确区分“AI 用于可持续发展”与“AI 自身是否可持续”两类问题：前者研究 AI 能帮助解决什么环境问题，后者研究构建和运行 AI 本身产生什么资源环境代价，两者不能彼此冲销。[28] AI 能够进入电网预测与调度、可再生能源功率预测、设备维护、建筑节能、工业控制、甲烷识别、天气气候分析、新材料与储能研究等领域。Rolnick 等系统梳理了机器学习在电力、交通、建筑、工业、气候预测和社会系统中的减排应用方向；IEA 2025 也把 AI 对能源优化、创新与安全的潜在作用作为报告另一半。[1,29]

12.1 正向作用必须从“技术潜力”走到“真实净结果”

一个算法在实验中减少 10% 能源，不等于整个行业净节能 10%。真正净结果还要经过部署率、成本、组织采用、基础设施更新、行为变化和反弹。AI 发现一种新电池材料也不等于它已经商业化；AI 提高电网预测也不等于电网具备足够输电和市场机制利用预测。环境收益同样必须遵守“能力→部署→现实结果”的完整因果链。

12.2 数据中心本身可以参与电力系统优化

Colangelo 等的 256-GPU 现场实验和 LBNL 关于可调数据中心的综述，为“把 AI 负荷从纯消耗者变成部分可调资源”提供了现实证据。[11–12] 如果未来越来越多训练、批处理或容忍时延的工作负载能够根据电网状态移时或迁移，数据中心可能减少高峰新增容量、增加可再生能源吸纳并降低局部约束。但需要明确：低时延在线推理、高安全任务、数据主权和网络限制可能显著降低这种灵活性。

12.3 正负两条链必须同时保留

表 5 AI 与环境的双向因果，而非单向宿命

路径	因果链
负向可能	AI 能力→更多计算→基础设施→能源/水/材料压力→生态与气候后果
正向可能	AI 能力→预测/控制/发现→能源与资源效率提高→环境压力下降
最终净结果	取决于两条链的规模、地点、时序、部署、反弹与替代，而不是由 AI“本质”决定

十三 | 多目标权衡：能源、水、碳、材料不能互相冒充

一个环境系统很少只有一个目标。空气冷却可能降低直接水耗却提高电力需求；蒸发冷却可能提高能源效率却增加耗水；建设大量储能提高可再生能源适配能力，却增加矿物与制造需求；延长硬件寿命减少隐含环境影响，却可能增加运行电耗。Chien 等 2026 年综述把水—碳权衡明确放进 AI 数据中心设计议程。[10]

13.1 不建立“绿色 AI 总分”

本文拒绝把能源、水、碳、土地、材料、生物多样性和社会影响压成一个不透明总分。原因不是这些指标不能比较，而是价值权重、地区边界和不可替代损害不能被技术人员悄然预设。专业环境分析更适合提供一张透明的多维证据表：每个指标的单位、范围、阈值、数据质量、地区意义和权衡是什么，再由经济、伦理、法律、政治和受影响者进入选择。

表 6 多维环境证据为什么需要分别保留

维度	典型指标	主要专业	禁止自动替代
电力/能量	TWh、MW、负荷曲线、PUE	能源系统与电网	不能替碳、成本或社会价值
水	取水/耗水、WUE、水足迹、流域压力	水资源/水文学	不能只用全国均值
气候	CO2e、平均/边际碳强度、生命周期排放	气候/能源/LCA	不能替生态多样性和水
材料	质量、关键性、供应集中、循环率	工业生态/资源环境	不能只看总吨位
生态/土地	栖息地、扰动、恢复、地方不可逆	生态学/环境科学	不能自动生成政治授权

十四 | 地球系统边界怎样进入 AI 研究，而不变成生态总分

地球系统科学提供一个重要尺度提醒：全球环境风险并不是无穷可替代的资源清单。气候、生物圈、淡水、土地、氮磷循环等过程相互作用，某些变化具有阈值、反馈和不可逆性。[22-25] AI 基础设施作为新增工业活动必须进入这些总体压力背景，而不能假定“数字行业”天然脱离地球系统。

14.1 行星边界不是项目许可线

Steffen 等明确讨论跨尺度作用和地区异质性；Richardson 等更新也强调多个边界之间的系统关系。[23-24] 因而，不能把“全球某边界被跨越”直接换算成“这个数据中心不能建”，也不能反过来因为一个项目占全球比例极小，就宣称其局部影响无关。项目层需要地点、生态、流域、电网与生命周期证据；全球框架负责说明累积后果和系统背景。

14.2 生态不可逆必须与可替代性分开

某些环境损失可通过修复、替代地点或技术改造缓解，另一些损失具有高度不可逆性或恢复周期远长于项目收益周期。专业研究应识别这种差异，而不是以“可以补偿”预先消除所有损害。F 类规范层面，本文主张不可逆生态损害不能因为 AI 带来的数字收益而在账面上消失；但是否接受特定权衡，仍需伦理、法律、政治与受影响者共同进入。

十五 | “智能丰裕”的现实条件：从数字可能到文明红利

“智能丰裕”最容易被误解为：只要模型推理越来越便宜，文明就会进入资源无限状态。事实上，智能服务的边际成本下降只解决了一部分约束。大规模智能还要经过电力、芯片、网络、建设、水、土地、材料、供应链和生态承载。稀缺不会因为一种能力变便宜而消失，而可能迁移到新的瓶颈。

15.1 第一重检验：效率是否真正快于总需求增长

若算法和硬件效率长期以快于总任务需求的速度提高，AI 的总体资源压力可能稳定甚至下降；若代理、多模态、视频与广泛自动化使需求扩张更快，总量仍会增长。这个问题必须用真实工作负载、硬件出货与能源数据回答，而不是依靠技术乐观或悲观立场。

15.2 第二重检验：增长是否能够进入电网而不挤压其它基本需求

当数据中心成为大负荷，真正的文明问题不是“它可不可以用电”，而是新增容量是否及时、可靠、低排放且成本责任清楚；是否使居民和工业承担不透明的基础设施成本；是否与关键公共服务、住房、电气化和工业发展形成冲突。这些后果首先需要能源、经济和政治共同重新进入，生态学不能独自裁决。

15.3 第三重检验：水、材料和生态是否具有可持续替代路径

电力问题解决后，水、芯片材料、关键矿物、土地和废弃物仍可能成为瓶颈。真正稳健的智能基础设施需要低水或适地冷却、资源循环、较长设备价值周期、多元供应链、生态适宜选址和可恢复设计。任何一种单一突破都不能替代整条物质链。

15.4 第四重检验：AI 的环境收益能否大于自身新增负担

如果 AI 显著改善能源系统、材料研发、气候适应和资源效率，其净环境结果可能远优于“只算数据中心负荷”的静态估计；反过来，如果 AI 主要创造高耗能低价值需求，而正向环境应用长期停留在技术演示，结果可能相反。本文把“净环境收益”保留为 C 类文明级待验证判断，不预先结算。

智能丰裕不是“算力无限”，而是智能扩张以后，真实世界仍有能力持续承载人的生活。

十六 | 跨学科交接：事实证明以后谁必须重新进入

本稿的专业责任在生态、能源、资源和地球系统处停止。资源事实一旦进入价格、权利、政治、安全、技术或文明目的，必须重新交接。交接不是把结论交给“上位学科”，而是防止专业越权。

表 7 生态／地球系统稿与相邻专业的判断交接

相邻专业	其独立责任	与本文的交接
AI／计算机科学	模型、芯片、算法、负荷、系统效率与技术可实现性	这些工程需求进入电网、水、材料与生命周期后由本文重新判断
系统科学	一般反馈、阈值、级联、韧性与锁定	本文提供具体地球、能源、水与生态事实；一般模式不能替真实测量
物理／信息／时空	物理能耗、热、因果与基本约束	基本物理约束不等于现实电网、资源和生态承载
经济／政治经济	价格、投资、所有权、收益与资源分配	生态稀缺与环境外部性进入经济后由其继续判断，价格不能反向证明生态安全
政治学	基础设施权力、国家能力、资源治理与公共授权	生态事实不能自动产生治理权
伦理／法学	环境权、代际责任、程序、公平与损害	本文不以生态指标决定最终权利冲突
国际关系／战略安全	关键矿产、能源与基础设施地缘依赖	供应集中如何进入战略稳定由战略专业重新判断
社会学／人类学	地方社区、土地、文化、接受度与生活后果	“地方影响”不能只由工程成本代表
生命科学	生物多样性、生物系统与生命影响	生态系统变化涉及生命机制时重新交接

16.1 五个明确不得越权

- 不得用“生态承载有限”直接推出某一种经济制度或增长制度必然正确。
- 不得用电、水或碳指标替人工智能与计算机科学规定模型架构和系统保证。
- 不得以生态阈值自动决定谁有权分配能源、水、土地和矿产。
- 不得把某一地区资源压力机械外推成全球 AI 不可持续。
- 不得用“环境更优”结算生存、权利、幸福与意义的全部文明价值。

十七 | 什么事实能够改变本文：最强反驳与可反驳研究议程

17.1 最强反驳一：AI 资源压力可能被效率提升迅速化解

这是本文必须认真接受的反方。数据中心能效历史上曾出现效率提升抵消需求增长的阶段；Masanet 等 2020 年就用更细的技术数据反驳了当时部分过高的全球数据中心能耗外推。[5] 如果未来算法、硬件、模型压缩和设施效率使单位资源强度下降速度长期显著快于服务需求扩张，本文关于“资源约束将成为重要文明变量”的强度应当明显降低。

17.2 最强反驳二：AI 可能成为净环境正收益技术

若 AI 在电网、能源创新、材料、工业过程、交通、建筑、气候适应和生态监测中的真实节能与减排，稳定、大规模、可归因地超过 AI 基础设施自身新增环境负担，那么把 AI 主要视为资源压力源将是错误的。RoInick 等与 IEA 提供了大量正向应用入口。[1,29] 这一反驳不要求否认数据中心物质性，而要求转向“净影响”研究。

17.3 最强反驳三：当前很多地方冲突可能主要是既有基础设施治理失灵

电网排队、住房成本、水权冲突、土地审批和社区不信任可能并非 AI 独有，而是既有规划、价格机制、基础设施投资和治理能力问题。若严格反事实研究表明，在相同新增负荷条件下，数据中心与其它工业负荷没有独立环境增量，本文应把相应问题交回能源、经济和治理研究，而不是用“AI”重新命名旧问题。

17.4 能够迫使本文收缩、重写或取消独立成篇的证据

表 8 本文的反证与撤销条件

证据方向	若未来出现什么	本文应怎样改变
单位资源强度	长期下降速度显著快于 AI 总需求，全球与主要地区总压力稳定下降	收缩“扩张压力”主张
电网承载	新增 AI 负荷可稳定通过灵活性、分布式资源与快速扩网消化，且不显著抬高成本/排放	降低电网瓶颈的重要性
水	直接和间接耗水在主要部署地区长期低且与当地水压力脱钩	收缩水资源章节
材料/废弃物	关键材料高度可替代、供应快速多元化，设备寿命延长且循环率显著提高	降低材料与 e-waste 判断强度
净环境收益	可归因的 AI 节能减排/生态收益稳定超过其新增环境负担	重写总体方向为净正效应研究
独立事实域	所有剩余问题均可被 AI 工程、系统科学和经济学无越权承担	重新审查本文独立成篇资格

17.5 下一轮可直接执行的专业研究

- 建立可复核的“任务—服务器—设施—电网—流域—生命周期”多尺度数据链，避免把单点指标外推为总体足迹。
- 开展跨地区真实工作负载水耗与电耗比较，公开服务器效率、利用率、冷却、电网和气候条件。
- 对 AI 训练、在线推理、批处理、视频生成和代理系统分别建立负荷曲线与灵活性边界。
- 把 2030 电力情景持续与真实芯片出货、设施并网、取消项目和利用率数据校准，而不是只累加公告容量。

- 构建包含芯片制造、服务器、建筑、备用系统、更新和废弃的 AI 生命周期清单，并公开数据缺口。
- 在高水压力地区开展流域尺度取水／耗水／季节性研究，而不是只报告企业全球年度水量。
- 比较硬件延寿、退居再用、模块化升级与新硬件节能之间的完整生命周期拐点。
- 检验 AI 用于电网和能源优化的真实净节能，区分算法潜力、实际部署和反弹。
- 把局部电网、流域和生态结果与社区、价格、就业及国际供应链事实交给相邻学科继续追踪。

结语 | 没有脱离地球的智能，只有更好地生活于地球的智能

AI 使人类第一次可以在极低边际成本下复制一部分过去昂贵的知识、分析和生成能力。这是真实的文明机会。但边际数字成本不是完整现实成本。任何一次计算都要经过芯片与服务器，任何一个数据中心都要落在土地、电网和流域中，任何一代硬件都来自材料并最终进入循环或废弃。技术越像“无形能力”，专业研究越有责任把它重新放回真实世界。

这并不导向反技术结论。恰恰相反，如果 AI 能够帮助电网更灵活、让能源与材料研发更快、使设备维护更精准、让资源利用效率更高，它可能扩大而不是缩小文明的生态承载。真正需要拒绝的只是两种自动推断：一是“数字所以无物质成本”；二是“有物质成本所以不应发展”。两者都跳过了专业证据。

生态学、地球系统与资源环境研究能做的，是把真实边界显影：哪里是年度总量问题，哪里是局部节点问题；什么已经被测量，什么只是情景；哪里存在水—碳—材料权衡；何种损失可修复，何种可能不可逆；单位效率提高以后总量是否真的下降；AI 为环境创造的收益是否真正实现。事实被看清以后，经济、政治、法律、伦理和文明方向才能重新进入。

任何关于未来智能丰裕的文明想象，都必须接受真实物质世界的检验。

文明的成熟，不是摆脱地球，而是在能力越来越强以后，仍然知道自己依赖什么、改变什么、消耗什么、修复什么，并让技术最终扩展人的生活，而不是把不可见的代价留给别的地区、别的生命或未来。

续读与共同研究入口

本文已经回答的是：AI 基础设施怎样进入电力、水、材料、土地、气候、废弃物与地球系统，哪些环境数字能够成立到哪里，哪些地方必须停止自动外推。继续往下，不是为了把读者留在同一套理论里，而是让问题沿真实因果继续进入。

向上 | 如果继续追问：这些物质边界为什么最终成为文明问题

进入《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，以及现实合流稿《AGI 时代的人工智能应用前景与风险研究——从能力扩展到文明后果》V1.2。前者说明为什么专业事实必须保持独立又必须重新接回共同因果链；后者继续追踪能源与算力怎样穿过科研、产业、治理与生活形成反馈。这里的文明尺度不是上位裁决，生态事实仍可以迫使整体理论收缩或重写。

向下 | 如果要把本文判断变成可验证现实

进入能源系统、电力工程、水资源、工业生态学、生命周期评价、材料流、关键矿产、区域环境监测和真实数据中心审计。最有价值的下一步不是再写一个“AI 有环境影响”的综述，而是形成可复核的多尺度数据：真实负荷、真实流域、真实硬件更新、真实 LCA、真实灵活性和真实生态结果。

向左 | 如果继续追问：资源事实进入其它因果链以后发生什么

进入 AI/计算机科学以重新判断算法、芯片和系统效率；进入经济学研究价格、投资和分配；进入政治学与法学研究基础设施授权、水权、土地和成本责任；进入国际关系/战略安全研究关键矿产、芯片、能源与基础设施依赖；进入社会学/人类学研究地方社区、接受度和分配后果；涉及生物多样性与生命系统时，由生命科学重新进入。

向右 | 如果不同意本文，应优先检验这些最强相反方向

重点进入“AI 环境压力被夸大”“高效计算将显著压低总需求”“核能/可再生能源/新型电网可消除电力瓶颈”“节水技术可迅速解除水约束”“反弹效应很弱”“关键材料可替代且供应能够快速扩张”“AI 对能源与气候的净贡献将显著为正”等竞争方向。相反性不是形式平衡；只要这些路径获得更强现实证据，本文就必须改变。

向前 | 哪些专业工作足以直接改变本文

新的实测数据、负结果、跨区域复现、流域研究、电网模拟与现场实验、材料供应数据、完整生命周期清单、设备退役和循环记录、净节能与反弹研究，都可以降低、删除或重写本文的判断。最有价值的参与，不是赞同，而是足以改变理论的专业工作。

文明未来链 | 如果这些问题被更好解决，人类可能生活在怎样的世界

如果希望进一步看见“技术能力受真实物质边界约束，同时又真正服务人的生活”进入工作、家庭、教育、社区、公共生活和长期文明以后可能形成怎样的世界，可进入《活的文明》三部曲以及五卷书中的未来文明主线。它们承担的是可审视、可修改、可拒绝的未来生活图景，不是本文生态与地球系统事实的证据。愿景不是证据；但研究最终仍要回答，我们希望这些事实共同把文明带向哪里。

文明跃迁五卷书：<https://www.civitas.top/documents/p02-five-001/> | 公众总入口：<https://www.civitas.top/documents/p01-entry-006/> | 完整文库：<https://www.civitas.top/library/>
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。

附录 A | 来源—主张边界登记

表 A1 主要来源与可支持范围

来源	本文采用的主张	明确停止点
IEA 2025/2026 [1-2]	全球数据中心电力、AI 驱动、能源供应、情景与瓶颈	可支持当前与 2030 情景；不可当作确定未来
LBNL 2026 [3]	美国 2030 数据中心用电底层模型与情景区间	只适用于其美国设备/设施假设；应随出货与利用率更新

来源	本文采用的主张	明确停止点
LBNL 水研究 [6]	工作负载耗水决定因素与>10,000 倍差异	支持“不能用单一每任务水耗常数”；不等于所有设施都会极端差异
Xiao 等 [9]	美国 AI 服务器水—碳情景与空间差异	属于模型情景，不是实测全国总量
Chien 等 [10]	AI 数据中心能源／水／隐含环境影响综合与权衡	综述支持多目标方法；具体比例依系统边界
Colangelo 等 [11]	256-GPU 现场负荷灵活性实验	证明某类工作负载可调；不能外推所有 AI 负荷
IEA 关键矿产 [13–14]	材料需求与供应集中	说明材料瓶颈可能性；不等于 AI 必然短缺
E-waste Monitor [16]	全球电子废弃物总量与回收背景	不是 AI 专属废弃物
Wang 等 [17]	生成式 AI e-waste 2020–2030 情景	B 类情景，不是 2030 确定事实
行星边界 [22–24]	全球地球系统风险框架	不能作为单项目自动许可阈值
IPCC/IPBES [25–26]	气候、生物多样性与系统风险背景	不替代 AI 具体归因
爱尔兰/弗吉尼亚/新加坡 [36–38]	地区真实承载案例	只证明地区差异与政策响应，不外推全球

附录 B | AI 资源环境压力测试矩阵

表 B1 资源环境压力测试矩阵

场景	首要环境维度	必须测量	重新进入专业	主要失败模式
大规模新建 AI 园区	电力	节点容量 / 输电 / 发电/峰值	区域能源系统	总电量够但接入延迟、成本或排放上升
高水压力地区冷却	水	直接耗水 / 间接电力水/季节流域压力	水文学/环境工程	WUE 低但电耗或碳上升；年度补水与枯水期错位
加速硬件更新	材料/e-waste	寿命、制造、退居、回收	工业生态/LCA	单位能效提高但隐含环境影响和废物上升
低碳能源配套	气候/土地/材料	实际电源、输电、储能、建设	能源/生态/材料	运行碳下降但土地/材料压力增加
AI 负荷灵活化	电网	任务延迟、迁移、SLA、数据位置	电力/AI 工程	技术可调但高安全/实时任务不可调
AI 用于节能	净环境收益	基线、归因、部署率、反弹	能源/经济/组织	实验节能无法进入真实系统或被新增需求抵消
关键矿产集中	材料/安全	需求份额、供应集中、替代、库存	资源/国际关系	材料总量小但供应冲击战略重要
生态敏感选址	土地/生物多样性	栖息地、恢复性、累积基础设施	生态/生命/社会	机房低影响但配套设施造成高生态成本

生成式人工智能使用声明

生成式人工智能在本文形成过程中可参与资料与文献线索整理、结构设计、部分文字起草与润色、交叉核查、引用线索核验和版式检查。AI 输出不作为经验数据，不作为权威来源，也不作为独立证据。研究问题、理论命题、规范判断、材料取舍、引用采用和最终定稿均由作者审定，作者承担最终学术责任，AI 不列为作者。

本文另加一项资源环境专门纪律：任何由生成式 AI 给出的能耗、水耗、碳排、材料需求、关键矿产份额、生态阈值或地球系统结论，均不得直接写入正文。所有定量材料必须回到原始研究、正式统计、标准、专业数据库

或可核验技术报告，并保留时间、地点、系统边界、方法和情景条件。正式投稿时，再按目标期刊要求补充具体工具、模型版本、使用日期与使用范围。

参考文献

- [1] International Energy Agency (IEA). (2025). Energy and AI. Paris: IEA. <https://www.iea.org/reports/energy-and-ai>
- [2] International Energy Agency (IEA). (2026). Key Questions on Energy and AI. Paris: IEA. <https://www.iea.org/reports/key-questions-on-energy-and-ai>
- [3] Smith, S. J., Hubbard, A., Newkirk, A., Ganeshalingam, M., Holecek, B., Sartor, D. A., Mills, M., & Shehabi, A. (2026). United States Data Center Energy Usage Report: 2025 Update. Lawrence Berkeley National Laboratory. <https://doi.org/10.71468/P1RP4F>
- [4] Shehabi, A., Smith, S. J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M. A. B., Holecek, B., Koomey, J. G., Masanet, E. R., & Sartor, D. A. (2024). 2024 United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory. <https://doi.org/10.71468/P1WC7Q>
- [5] Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984–986. <https://doi.org/10.1126/science.aba3758>
- [6] Lei, N., Lu, J., Shehabi, A., & Masanet, E. R. (2025). The water use of data center workloads: A review and assessment of key determinants. *Resources, Conservation and Recycling*, 219, 108310. <https://doi.org/10.1016/j.resconrec.2025.108310>
- [7] Mytton, D. (2021). Data centre water consumption. *npj Clean Water*, 4, 11. <https://doi.org/10.1038/s41545-021-00101-w>
- [8] Siddik, M. A. B., Shehabi, A., & Marston, L. (2021). The environmental footprint of data centers in the United States. *Environmental Research Letters*, 16, 064017. <https://doi.org/10.1088/1748-9326/abfba1>
- [9] Xiao, T., Nerini, F. F., Matthews, H. D., Tavoni, M., & You, F. (2025). Environmental impact and net-zero pathways for sustainable artificial intelligence servers in the USA. *Nature Sustainability*, 8, 1541–1553. <https://doi.org/10.1038/s41893-025-01681-y>
- [10] Chien, A. A., Gupta, U., Ren, S., Sriraman, A., & Tomlinson, B. (2026). Strategies and design for increasing AI sustainability. *Nature Reviews Clean Technology*. <https://doi.org/10.1038/s44359-026-00195-w>
- [11] Colangelo, P., Coskun, A. K., Megrue, J., et al. (2026). AI data centres as grid-interactive assets. *Nature Energy*, 11, 254–261. <https://doi.org/10.1038/s41560-025-01927-1>
- [12] Granderson, J., Hoffman, I. M., Holecek, B., Crowe, E., Smith, S. J., & Mims Frick, N. (2026). Integrating AI Data Centers with the Power Grid. *The Bridge*. <https://doi.org/10.20357/B7X61S>
- [13] International Energy Agency (IEA). (2025). Global Critical Minerals Outlook 2025. Paris: IEA. <https://www.iea.org/reports/global-critical-minerals-outlook-2025>
- [14] International Energy Agency (IEA). (2026). Global Critical Minerals Outlook 2026. Paris: IEA. <https://www.iea.org/reports/global-critical-minerals-outlook-2026>
- [15] United Nations Environment Programme, International Resource Panel. (2024). Global Resources Outlook 2024: Bend the Trend – Pathways to a Liveable Planet as Resource Use Spikes. Nairobi: UNEP. <https://www.unep.org/resources/Global-Resource-Outlook-2024>
- [16] International Telecommunication Union (ITU) & United Nations Institute for Training and Research (UNITAR). (2024). The Global E-waste Monitor 2024. Geneva/Bonn. https://www.itu.int/hub/publication/d-gen-e_waste-01-2024/
- [17] Wang, P., Zhang, L.-Y., Tzachor, A., & Chen, W.-Q. (2024). E-waste challenges of generative artificial intelligence. *Nature Computational Science*, 4, 818–823. <https://doi.org/10.1038/s43588-024-00712-6>
- [18] Plociennik, C., Watjanatepin, P., Van Acker, K., & Ruskowski, M. (2025). Life Cycle Assessment of Artificial Intelligence Applications: Research Gaps and Opportunities. *Procedia CIRP*, 135, 924–929. <https://doi.org/10.1016/j.procir.2025.01.079>
- [19] International Organization for Standardization (ISO). (2006). ISO 14040: Environmental management — Life cycle assessment — Principles and framework.
- [20] International Organization for Standardization (ISO). (2006). ISO 14044: Environmental management — Life cycle assessment — Requirements and guidelines.
- [21] International Organization for Standardization (ISO). (2014). ISO 14046: Environmental management — Water footprint — Principles, requirements and guidelines.
- [22] Rockström, J., Steffen, W., Noone, K., et al. (2009). A safe operating space for humanity. *Nature*, 461, 472–475. <https://doi.org/10.1038/461472a>
- [23] Steffen, W., Richardson, K., Rockström, J., et al. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science*, 347(6223), 1259855. <https://doi.org/10.1126/science.1259855>
- [24] Richardson, K., Steffen, W., Lucht, W., et al. (2023). Earth beyond six of nine planetary boundaries. *Science Advances*, 9(37), eadh2458. <https://doi.org/10.1126/sciadv.adh2458>
- [25] IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report. Geneva: IPCC. <https://doi.org/10.59327/IPCC/AR6-9789291691647>
- [26] IPBES. (2019). Global Assessment Report on Biodiversity and Ecosystem Services. Bonn: Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. <https://doi.org/10.5281/zenodo.3831673>
- [27] Sorrell, S. (2009). Jevons' Paradox revisited: The evidence for backfire from improved energy efficiency. *Energy Policy*, 37(4), 1456–1469. <https://doi.org/10.1016/j.enpol.2008.12.003>
- [28] van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1, 213–218. <https://doi.org/10.1007/s43681-021-00043-6>

- [29] Rolnick, D., Donti, P. L., Kaack, L. H., et al. (2022). Tackling Climate Change with Machine Learning. *ACM Computing Surveys*, 55(2), Article 42. <https://doi.org/10.1145/3485128>
- [30] Nuss, P., & Eckelman, M. J. (2014). Life cycle assessment of metals: a scientific synthesis. *PLOS ONE*, 9(7), e101298. <https://doi.org/10.1371/journal.pone.0101298>
- [31] Frosch, R. A., & Gallopoulos, N. E. (1989). Strategies for manufacturing. *Scientific American*, 261(3), 144–152.
- [32] International Energy Agency (IEA). (2026). Critical Minerals Dataset, 2026 edition. Paris: IEA. <https://www.iea.org/data-and-statistics/data-product/critical-minerals-dataset>
- [33] OECD. (2022). Measuring the environmental impacts of artificial intelligence compute and applications: The AI footprint. *OECD Digital Economy Papers*, No. 341. <https://doi.org/10.1787/7babf571-en>
- [34] ISO/IEC. (2026). ISO/IEC 30134-2:2026 Information technology — Data centres key performance indicators — Part 2: Power usage effectiveness (PUE).
- [35] ISO/IEC. (2022). ISO/IEC 30134-9:2022 Information technology — Data centres key performance indicators — Part 9: Water usage effectiveness (WUE).
- [36] Central Statistics Office Ireland. (2026). Data Centres Metered Electricity Consumption 2025. Statistical Release, 7 July 2026. <https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2025/>
- [37] Joint Legislative Audit and Review Commission (JLARC), Commonwealth of Virginia. (2024). Data Centers in Virginia. Richmond, VA. <https://jlarc.virginia.gov/landing-2024-data-centers-in-virginia.asp>
- [38] Infocomm Media Development Authority (IMDA), Singapore. (2024). Green Data Centre Roadmap. <https://www.imda.gov.sg/-/media/imda/files/news-and-events/media-room/media-releases/2024/05/green-dc-roadmap.pdf>

版本说明

V1.1 在 V1.0 首个完整专业版本基础上完成独立专业审查后的必要受控修订：精确化 256-GPU 电网互动现场演示；将 IPBES 生物多样性证据显式接入土地／生态选址章节；分开 2025 年数据中心需求情景与 2026 年关键矿产供应集中度更新；修正 Xiao 等 2025 论文 DOI 并补充动态／正式来源稳定入口。理论主体、研究对象、A／B／C／F 身份、成熟专业遗产、跨学科交接、不得越权事项、反证条件、上下左右＋向前＋文明未来链和生成式 AI 使用声明均保持不变。本次跨稿一致性同版回补仅同步现实合流稿版本与统一公开入口，不改变生态专业判断。V1.1 为当前专业研究基线。