

AGI 十年内会出现吗

制品 · 持续 · 预判

AGI 十年内会被宣布，但绝不会被公认

报告出具时间：2026/6/29

报告来源：See The Essence

目录

一、全貌	3
二、核心机制	4
制造：定义空白是构造空间	4
传播：三次放大	4
后果：资源重配与注意力偏移	4
遗产：期望曲线的上行段	4
三、深层洞察	5
定义权空缺是权力的操作系统，不是等待修补的 bug	5
物理定律是实验室演示与普惠基础设施之间真正的距离	5
治理脆弱性是最被低估的系统刹车	5
四、走向与可能	6
最可能的走向：宣布发生，共识不发生	6
被忽视的约束：基础设施的沉默投票	6
最意外的可能：开源生态爆破式涌现	6
五、如何思考这件事	7
参考资料	8

AGI 十年内会不会出现——这个问题本身比它的答案更有信息量。提出它的人主要是一群正在烧钱做这件事的机构，他们的回答在功能上维持着估值叙事。而“AGI 出现了”这句话在人类历史上从未有过公认的判定标准。即使实验室跑通了，让它在物理世界安全运转需要越过的不只是算法，是一组物理、工程和治理的硬约束。

一、全貌

“AGI 十年内会出现”这个判断, 从 2020 年代初期开始密集出现。2023 年 GPT-4 发布后, Sam Altman 提到“未来十年内”, 部分前沿实验室给出 2026-2030 年的区间[1]。这个时间窗口恰好与资本涌入同步——流向生成式 AI 的风险投资在一年内翻了三倍以上[2]。

技术预测从来不只是关于技术。当一家公司估值取决于投资者相信它正在建造通向 AGI 的桥梁, “AGI 很快会来”就不是科学判断, 而是组织现在的话语实践——吸引顶尖工程师、说服主权基金押注、让监管者紧张。任何一个声称在冲刺这件事的机构, 都有充分理由让外界相信终点更近、自己更领先。

但这带来了真正的问题: 我们至今没有一种方法能区分“真的通用智能”和“好到让人愿意相信的模拟”。“十年内 AGI 会出现”和“十年内有人说 AGI 出现了”是两个完全不同的问题——前者需要共识, 后者只需要一次发布会。

二、核心机制

作为话语制品的“AGI 十年内会出现”，从制造到传播再到后果，遵循着一套可追踪的逻辑。

制造：定义空白是构造空间

AGI 的定义本身就是最脆弱的构件。百度百科说它是“能够像人类一样理解、学习和应用知识完成各种任务的智能系统”[3]，但经不起推敲——“像人类一样”是哪个认知水平？“理解”是通过测试还是需要意识？“各种任务”是谁来抽样？

哲学界对“什么是通用智能”争论了几十年，没有收敛迹象[4]。复杂系统理论提供了一个关键视角：通用能力是涌现属性，不是专项能力的加总，不存在天然的“开关时刻”。这就给了制造方巨大的话语空间——如果标准由研发机构定义，AGI 可以在任何需要宣布的季度被宣布。2018 年就有团队在特定基准集上宣称“类人水平”，但标准没获跨社群认可，也就没人在意。这个模式不因为参数从千亿变万亿就失效。

传播：三次放大

这个叙事从实验室进入公众认知，经历了三个放大节点。

GPT-4 的发布让大量非技术用户第一次体验到流畅对话和跨领域推理，形成直观印象——只是少了一些什么。资本市场随后吸收叙事，一级市场需要 AGI 论证数千亿估值[5]，二级市场用“AGI 前夜”解释资本开支飙升。最关键的是第三跳：当 AI 安全机构开始严肃讨论“AGI 十年内出现”的应对方案，公众的逻辑变成“最谨慎的人都这么说了，大概是真快了”。但安全社区推演最坏情境是其职责所在，这不意味着他们认为这个情境最可能——传播中，情景推演被转写成预测，预测再变成既定事实。

后果：资源重配与注意力偏移

人才向头部实验室集中，资本以数千亿美元涌入，政府把 AGI 当作地缘竞争核心制定出口管制和补贴政策。更隐蔽的后果是，当“直接冲刺 AGI”成为优先级，那些五年内能解决实际问题的渐进改进——可解释性、可审计性、跨文化对齐——被系统性地搁置。

遗产：期望曲线的上行段

核聚变在 1950 年代预测“五十年内商用”，七十多年后仍是“再有三十年”。AI 领域经历了两次寒冬：1970 年代符号学派在自然语言理解上承诺过度，1980 年代末专家系统维护成本击溃商用预期。每次突破启动承诺周期，资源涌入；承诺无法在融资周期兑现，幻灭到来。“十年内 AGI 会出现”最可能的遗产不是 AGI 本身，而是它制造的期望曲线上涨段，以及回调时的新一轮幻灭。

三、深层洞察

定义权空缺是权力的操作系统，不是等待修补的 bug

通常默认假设是“暂时还没找到公认定义，早晚会有”。但过去七十年的辩论表明，“缺少公认定义”可能是“AGI 宣称”这个社会行为的构成性条件。

如果存在公认的、不可作弊的通用智能评估标准，研发机构就失去了定义里程碑的能力——他们会面临不可接受的被独立验证推翻的风险。因此 AGI 定义之争不会在十年内解决。十年后最可能的局面是：某些系统在部分指标上超越人类的泛化能力，但另一阵营坚定指出它仍缺常识、因果理解或具身经验。双方各有证据，无人能说服对方。这就是“话语冲突”的核心事件本身。

物理定律是实验室演示与普惠基础设施之间真正的距离

大模型性能提升与计算量呈幂律关系，每获得一个单位性能增量需指数级算力投入。目前前沿模型单次推理能耗已达每用户数十千瓦时级别——相当于四口之家数天用电量[6]。

这有两层硬约束。第一是芯片工艺：硅基芯片逼近原子尺度，3 纳米以下每走一代成本倍增、良率下降，而 AGI 所需算力远在可见工艺曲线之外。第二是电网：AGI 级别系统部署到数十亿用户，总能耗将是全球发电量的显著比例，当前没有国家的基础设施规划纳入这种负载。即使算法突破让 AGI 在实验室跑起来，把它放大为社会范围的通用智能基础设施，需要的时间尺度远不止十年。

治理脆弱性是最被低估的系统刹车

2023 年底 OpenAI 的治理危机暴露了一个致命结构：立志构建 AGI 的非营利机构，董事会合法解雇 CEO，CEO 通过员工集体辞职威胁和投资者施压在几天内夺回。这意味着在那家机构，最高领导者在实质分歧中不受任何约束。

这不是 OpenAI 一家的问题。前沿实验室面临同一个结构性矛盾：吸引资本和人才需要接近创业公司的股权结构和高风险偏好，承担“创造 AGI”的全球后果需要比核武器研发更严格的集体约束。这两个要求指向完全相反的组织形式。在市值驱动的逻辑和地缘军备竞赛逻辑双重奖励下，组织层面几乎没有能力在到达临界点前踩下必要的刹车。

四、走向与可能

最可能的走向：宣布发生，共识不发生

十年内至少有一家前沿实验室将宣布其系统达到自定 AGI 阈值——发布会、精选基准测试、媒体演示，投资人和股价跟进。但这个宣布无法获得跨社群公认。另一阵营的科学家会列出系统在分布外因果推理、稳定自我模型、跨文化深层偏见、无人工干预持续学习上的崩溃证据。社交媒体成为双方产出证据的主战场，终极结论遥不可及。政府面对“已出现的 AGI”和完全缺乏共识的技术定义，立法将滞后，真空期内过度乐观者继续加注，安全担忧者持续焦虑。

前沿实验室的决策层和独立研究者及安全社群，双方动力是结构性的、对撞的，没有和解机制。

被忽视的约束：基础设施的沉默投票

假设算法出现突破，系统可以在一个任务上泛化到任意新任务。让它从 100 个用户扩展到 10 亿用户，在真实世界的不稳定电力、湿热水冷数据中心、参差不齐的终端设备中稳定运行，需要数万亿美元电网投资、新一代芯片架构量产、跨域数据治理的法律框架。芯片从设计到量产需五到七年，电网改造在发达国家是十年一个代际，全球数据治理框架乐观估计需二十年以上。即使原型真的造出来了，可部署的 AGI 将在物理瓶颈前卡住多年。这个约束不能用更聪明的工程绕过，它是物理学。

最意外的可能：开源生态爆破式涌现

过去两年开源模型追赶速度超过多数分析师预测。如果这个趋势持续，在某个节点当开源模型在足够多维度对齐甚至超过闭源水平，AGI 就不再被任何单一机构掌控——数万开发者发现自己手里的工具能做出以前不敢想的事，通用能力不再是宣称而是扩散的事实。这个情景的阻碍也很明确：主权国家限制开源发布、云服务商收紧算力准入、训练成本指数增长超出社区承受能力，任一项生效都可能封堵这条路。

五、如何思考这件事

换掉那个只能得到宣传口号的问题，持续追问三个能校准判断的问题。

如果有机构宣布 AGI 已实现，它通不过哪一项测试？ 分析宣称看的是失败项，不是通过项。如果系统通过 100 项基准，在第 101 项——儿童能做的物理因果推理——上系统性崩溃，宣称质量取决于第 101 项的解释力。持续追问“失败在哪”是防止被基准过拟合蒙蔽的最有效方法。

支撑进展的资源曲线是否延续？ 如果前沿模型单次训练成本增长快于可用资本和算力增长，“指数趋势”将在某一点自然弯曲。芯片成本曲线和电网投资曲线的斜率比论文引用曲线更值得关注——它慢，但不撒谎。

治理结构有没有出现能约束核心决策者的新机制？ 当前局面下，前沿实验室 CEO 有实质上不受约束的权力决定声称什么、释放什么。在出现比“董事会解雇-员工反叛-CEO 复辟”更有效的约束之前，不要相信任何“我们会负责任地释放”的承诺——盯着组织的行为，不是声明。

最容易被普遍接受但可能错的假设是这个：智能是一根可以无限攀升的单一标尺。复杂系统研究和认知科学指出，泛化能力、工具使用、社会认知、因果推理可能是多个正交维度。一个系统可以在三个维度超人，在第四个维度还停留在蟑螂水平——而那块短板在真实开放世界中恰好最致命。如果这个假设是错的，那么“十年内 AGI”就相当于“我们正以每小时一百公里驶向悬崖，预计十分钟到达”。到达确实可能。到达的方式可能和所有人想象的不一样。

参考资料

- [1] [新闻报道] 腾讯新闻, “Altman 再谈 AGI: 十年内 AGI 将会到来”, 2026 年 6 月发布
- [2] [机构报告] 36 氪, “2024 年 AI 投融资趋势报告”, 2025 年发布
- [3] [百科] 百度百科, “通用人工智能 (AGI)” 词条
- [4] [学术论文/博客] 知乎专栏, “通用人工智能的哲学边界: 为什么我们无法定义‘理解’”
- [5] [新闻报道] 搜狐, “AGI 前夜的资本狂潮”, 2025 年发布
- [6] [技术分析] 腾讯云开发者社区, “大模型推理能耗分析”, 2024 年发布