

AGI 的工程化定义：以复杂项目闭环能力为判断标准

版本：V1.1

作者：刘明

作者身份：独立研究者，北京，中国

ORCID：0009-0001-2816-538X

电子邮箱：erran@zymobile.net

日期：2026 年 9 月 3 日

摘要

人工通用智能（Artificial General Intelligence, AGI）长期缺少稳定且可操作的定义。既有定义分别强调拟人行为、主观心智、广泛认知能力、人类水平表现、经济价值或社会影响，却仍难以回答一个工程问题：AI 何时从高能力工具转变为能够独立承担复杂生产责任的系统？本文将判断单位从局部能力和单项任务移向复杂项目闭环，提出工程化 AGI 的工作定义：在目标、资源、安全边界与初始上下文给定，且验收边界能够独立建立的情况下，AI 系统能够在无持续人工驱动的情况下，稳定完成多数原本需要普通人类团队长期协作的典型复杂项目，并提供可验证的交付结果。本文进一步明确，该定义主要覆盖路径可能未知但能够由外部目标与评价条件闭合的复杂项目；它不要求系统在尚无稳定问题表述或评价标准的开放问题上持续产生超越人类前沿的原创贡献。本文的结构增量主张是：AGI 的关键判断单位应从局部能力提升到可闭合复杂项目中的责任连续性，并与开放前沿创新能力明确区分。本文的方向价值主张是：这一边界使 AGI 能够通过项目覆盖、人工介入、交付可靠性和独立验收证据被连续测量，为 AI 工程、知识工作自动化和生产力判断提供可比较、可验证并可逐步逼近的系统级目标。该定义不试图解决意识、主观体验等哲学问题。

关键词：人工通用智能；工程化 AGI；复杂项目；可闭合项目；项目闭环；自主系统

1. 引言

人工通用智能长期被视为人工智能发展的核心目标，但“AGI 是否已经接近”仍缺少统一的判断基础。图灵将机器智能问题转化为可观察的行为测试[1]；Searle 对程序是否足以构成理解和心智提出了哲学质疑[2]；Legg 与 Hutter 从智能体在广泛环境中实现目标的能力出发，提出了形式化的通用智能定义[3]。此后，人类水平表现、经济工作替代、能力广度与深度、认知结构以及社会转型影响，分别成为 AGI 定义和测量的重要路径[4-9]。

这些路径各自回答了重要问题，却没有完全解决同一个工程缺口：一个在大量局部任务中表现出色的 AI，为什么仍可能无法接管真实复杂项目？局部任务主要检验特定能力；复杂项目还要求系统长期维持目标，处理任务依赖、过程变化、错误恢复、结果验收和后续维护。只要这些工作仍主要依赖人类持续推动，AI 就仍是高能力工具，而不是项目责任的承担者。

本文据此提出一个面向工程实践的 AGI 工作定义。其目的不是取代所有 AGI 定义，而是补充一个更接近真实生产力形成过程的判断层：AI 能否把已有能力持续组织为可靠的复杂项目交付。

2. 工程化 AGI 的工作定义

2.1 核心定义

本文提出：

工程化人工通用智能，是指 AI 系统在目标、可用资源、安全边界与初始上下文给定，且验收边界能够独立建立的情况下，能够在无持续人工驱动的情况下，稳定完成多数原本需要普通人类团队长期协作的典型复杂项目，并提供可验证交付结果的系统级能力状态。

简化表达为：

工程化 AGI，是 AI 无人化完成原本需要普通人类团队长期协作的复杂项目的能力。

简化表达适合公共传播，正式定义用于研究和评估。这里的核心不是 AI 是否像人，而是项目推进所需的主要脑力劳动、过程治理和交付判断，能否由 AI 稳定承担。

2.2 关键概念

AI 系统不是单一基础模型，而是模型与记忆、工具、执行环境及必要治理机制共同形成的可运行整体。同一模型在不同系统配置下可以呈现显著不同的项目能力，因此工程化 AGI 是系统级属性。

复杂项目不是简单的长任务，而是包含多层依赖、阶段性目标、过程反馈和整体交付要求的工作单元。它通常需要在不完全信息和变化条件下持续规划、执行、验证与修正。对于没有明确终点的持续运营型工作，可以将具有明确目标、周期和验收条件的阶段性运营单元作为项目闭环的判断单位。

可闭合项目不等于解决路径预先已知，也不等于所有局部步骤都有唯一正确答案。它是指项目目标、主要约束与验收边界能够在系统给出最终方案之前，由外部独立建立或在项目过程中稳定下来。软件开发、实验复现、产品实施和许多分析项目都可能包含不完全信息、局部探索与方案创新，但只要其最终交付仍能依据相对独立的条件验收，就属于本文主要讨论的可闭合范围。

普通人类团队是项目难度的参照基线，指具备相应职业能力、使用常规工具和组织流程的专业团队，而不是单个普通人，也不是由顶尖专家组成的特殊团队。具体评价中应声明所采用的人类团队基线及其典型资源、周期和能力范围。

无持续人工驱动不等于完全没有人参与。人类仍可设定目标、配置资源、规定安全边界并承担最终授权；但 AI 不再依赖人类持续拆分任务、纠正方向、恢复上下文、修复执行过程或代替其完成常规验收。

稳定完成不等于一次成功或形式上产出完整。它要求系统在同类项目中具有可重复的交付能力，结果能够接受独立验收，并且没有通过缩减需求、隐藏失败或把关键工作转移给人类获得表面成功。

2.3 项目闭环

本文所说的项目闭环，是指系统能够从给定目标出发，持续组织项目所需的主要认知与治理活动，形成可验收结果，并在出现错误和合理变化时维持交付能力。闭环关注的是责任连续性，而不是 AI 是否亲自执行每一个局部动作。

因此，AI 可以通过软件接口、自动化设备、机器人、专业服务或受指导的人类完成部分现实操作。外部人类可以承担授权、物理执行或明确界定的服务接口，但不得用于补充系统自身缺失的项目规划、关键判断、过程纠错或常规质量控制能力。只要项目所需的核心判断、协调和质量控制主要由 AI 承担，这些外部接口就不改变项目闭环的性质。

项目闭环允许系统在执行中生成新方案、新表示和局部新结构。工程化 AGI 与更高层开放智能的区别，不在于前者完全不产生新结构，而在于这些结构仍服务于一个能够从外部闭合的项目目标，其价值可以由独立于最终方案的交付条件进行判断。

3. 为什么复杂项目闭环是关键判断单位

3.1 局部智能不直接等于生产力跃迁

当前 AI 已经能够写代码、生成文档、分析材料、辅助设计并完成大量独立任务。但在多数真实场景中，人类仍负责提出具体问题、拆解任务、协调依赖、判断结果、纠正错误并承担维护责任。AI 提高了局部效率，却尚未普遍改变项目责任结构。

工程化 AGI 的转折点在于：AI 不再只是被调用的能力模块，而开始承担项目推进责任。人类从持续驱动者转变为目标设定者、边界制定者、风险审计者和最终授权者。

这一变化比局部任务性能更直接地连接社会生产力。软件系统、科研计划、制造流程、教育平台和组织运营等重要产出，本质上都依赖复杂项目能力。当项目执行、协调和维护成本显著下降后，大量过去因组织成本和试错成本过高而无法启动的项目才可能变得可行。

本文选择团队而非单个人类作为参照，并非主张 AGI 必须在认知能力上等于多人智能之和，而是因为现代复杂生产的自然责任单元通常由团队承担。工程化定义所测量的，是 AI 系统能否接管这一生产责任单元，而不是它是否在个体认知测验上达到某一人类水平。

3.2 项目能力不是局部能力的简单相加

复杂项目具有路径依赖。早期需求理解、技术选择或资源配置的偏差，可能在后续阶段被放大；一个局部正确的结果，也可能破坏项目整体一致性。因此，项目闭环至少要求系统具备跨阶段目标保持和全局协调能力。

这些能力可以由外部记忆、审计程序和工具系统共同提供，不要求全部存在于单个模型内部。但无论采用何种实现方式，最终判断都应落在系统能否持续形成可靠交付，而不是系统在多少个孤立任务上取得高分。

近年来，以人类基线完成时间标定 AI 可可靠完成任务规模的研究，已经开始把评估从静态题目推向更长时程的工作[10]。PaperBench 等评测则进一步要求 AI Agent 从零开始理解研究论文、开发代码并执行实验，开始接近端到端复杂工作的评价形式[11]。这些方向说明任务长度、可靠性和端到端执行具有独立价值，但真实复杂项目还包含多层依赖、过程变化、维护责任与外部协作，因此不能仅由任务时长或单一领域的端到端任务概括。

4. 与既有 AGI 定义的关系

既有 AGI 定义选择了不同的判断单位。它们与本文定义并非简单竞争关系，而是分别描述了人工智能的不同层面。

定义路径	主要关注点	对工程化判断的意义与不足
行为拟人路径	机器是否表现得像人[1]	提供外部行为标准，但拟人表现不等于项目交付能力
心智与理解路径	机器是否真正理解或具有心智[2]	处理重要哲学问题，但难以成为工程验收标准
形式化通用智能路径	智能体在广泛环境中实现目标的能力[3]	理论抽象更一般，但与真实项目证据仍有距离
人类水平与认知路径	能力广度、深度及相对人类的表现[4,6-8]	可以描述能力结构，却不能单独证明长期责任连续性
经济工作路径	在多数经济价值工作上超过人类[5]	将 AGI 与生产力联系起来，但判断范围较宏观
社会影响路径	是否引发工业革命级社会转型[9]	适合描述后果，但通常只能事后确认

认知能力测量可以说明系统拥有哪些基础能力，任务评估可以说明系统在给定条件下能完成什么，项目闭环评价则检验这些能力能否被持续组织为可靠交付。三者可以互相提供证据，但不能彼此替代。

工程化定义与经济工作定义最为接近。能够广泛接管复杂项目的系统，很可能产生显著经济影响；但经济结果还受到成本、制度、监管和部署速度影响。工程化定义把判断位置前移到系统能力本身，不必等到宏观影响已经发生后再确认。

5. 定义边界

为避免对定义作过度外推，本文集中说明以下边界。

5.1 工程化工作定义与人类责任边界

本文定义适合讨论 AI 工程、知识工作自动化、复杂项目交付和组织生产替代，不直接回答机器是否具有意识、主观体验或真正理解。一个系统可以满足本文定义，而其心智地位仍然存在争议；反之，即使某种机器意识成立，也不必然具有复杂项目能力。

“承担项目责任”在本文中是功能性表述，指 AI 能够维持目标、组织执行、暴露问题、推动恢复并形成可验收结果。人类仍然负责目标设定、资源授权、安全边界、高风险伦理判断和最终社会责任。工程化 AGI 描述的是项目执行与治理责任的转移，不意味着 AI 自动成为法律或道德责任主体。

5.2 工程化 AGI 不以完整具身能力为前提

现实中的复杂项目本来就依赖团队、工具和执行接口，不由某一个人完成全部物理操作。AI 是否拥有类人身体，不应成为所有领域共同的 AGI 前提。机器人和专用自动化设备会扩大可承担项目的范围，但不是本文定义成立的必要条件。

5.3 单一领域的项目自治不等于通用智能

一个系统可能在软件开发、药物筛选或制造调度等单一领域形成较强项目闭环，但这只能证明领域能力。AGI 中的“通用”要求其能力覆盖具有足够广度，不能由少量特制项目、单一垂直系统或事后选择的成功案例支持。

这里的“多数典型复杂项目”只有相对于预先声明的项目集合或代表性抽样分布才具有可检验意义。任何关于系统已达到工程化 AGI 的经验性主张，都应说明项目范围、抽样原则及领域覆盖，避免以事后筛选的成功案例替代通用能力证据。

5.4 稳定完成、可验证交付与可闭合边界

工程系统不存在绝对无误。本文所说的稳定完成，是指在明确项目范围、资源条件、验收标准和风险水平下，系统在同类项目中达到足以承担交付责任的可重复可靠性，而不是要求所有项目必然成功。不同领域可以具有不同门槛，高风险领域也必须附加更严格的安全要求。

“可验证交付”也不要求所有项目都存在唯一客观真值。对于软件、制造等结果可直接测试的领域，可以采用明确验收标准；对于实验复现、产品实施、方案比较和战略分析等路径开放但结果可闭合的工作，则应依据预先设定或能够独立建立的评价条件进行外部验收。其核心要求是交付结果不能仅由系统自身宣称成功。

如果一个问题的关键困难恰恰在于发现什么值得研究、重新表述问题、提出新的评价维度，或者判断一个尚无既定标准的新结构是否值得保留，那么它尚未被充分闭合。工程化 AGI 可以参与此类工作，也可能偶尔产生重要贡献，但本文定义不以系统能够在这类问题上稳定超过人类认知前沿为成立条件。

5.5 本文不预设唯一评测标准

不同领域的项目结构、风险和验收方式差异很大，不宜过早规定统一指标或固定阈值。项目覆盖范围、任务依赖规模、人工介入程度、交付可靠性、失败恢复、维护连续性、完成周期和资源成本，都是可能的测量方向；具体评价体系需要在各领域通过后续研究建立。

但缺少统一指标并不意味着评价可以缺少操作条件。任何具体比较或工程化 AGI 主张，至少应声明：所采用的项目集合或抽样分布、普通人类团队基线、允许使用的资源与时间条件、人工介入规则、验收条件以及成功判据。只有在这些条件明确后，“多数”“典型”“稳定完成”和“无人化”才具有可比较的经验含义。

5.6 工程化 AGI 与 ASI 的边界

部分既有框架把人工超级智能（Artificial Superintelligence, ASI）放在广泛人类相对性能的最高层，例如要求系统在广泛任务上超过所有人类[6]。开放式智能研究则进一步指出，持续产生对人类观察者而言新颖且可学习的成果，是人工超级智能的重要条件[12]。本文不在此完整定义 ASI，但需要明确工程化 AGI 的上边界：工程化 AGI 所要求的是对可闭合复杂项目的广泛自主交付能力，而不是在真正开放的问题上持续产生超越人类前沿的方向、抽象或评价框架。

因此，从当前 AI 到工程化 AGI，核心变化是复杂项目执行与治理责任由人类持续驱动转向 AI 自主闭环；从工程化 AGI 到 ASI，还需要进一步讨论开放问题中的方向发现、结构生成与价值判断能力。这两个变化不能仅用同一组局部任务分数概括。

6. 工程化定义如何使 AGI 成为可逼近目标

工程化定义的价值，不在于立即给出一个统一分数，而在于把 AGI 从模糊想象转化为可以持续观察的能力方向。研究者不再只需争论系统是否“像人”或是否已经拥有某种一般智能，而可以追问：它能承担多大规模、多长周期和多广范围的项目？项目对人工推动的依赖是否下降？交付可靠性、失败恢复和长期维护能力是否改善？

这些问题能够形成连续证据。即使尚无系统达到最终门槛，也可以在明确项目分布、资源条件和人工基线后，比较不同系统可承担的项目范围、复杂度和自治程度。单点能力进步由此被放入更完整的生产结构中理解，而 AGI 也不再只是一个等待宣布的二元事件。

这一视角还可以解释当前 AI 发展中的一个表面矛盾：局部能力快速提升，并未同等幅度地转化为企业组织变化和社会生产力跃迁。缺失的环节不是又一种局部技能，而是把多种能力持续组织为项目闭环的系统能力。

7. 讨论：从能力评测到责任连续性

现有能力基准仍然是工程化 AGI 评估的重要组成部分，但需要被放入更高层的证据结构中：局部能力测试说明系统具备哪些能力，任务评估说明系统在给定条件下能够完成什么，项目评估则检验这些能力能否跨阶段持续组织为可维护交付。

因此，复杂项目闭环并不是用单一新基准替代既有评测，而是改变最终判断单位。随着系统能力提高，评估重点需要逐步从单次能力展示转向项目范围、自治程度、失败恢复、长期维护和责任连续性。工程化 AGI 的关键标志，不只是系统在某个时刻表现得聪明，而是它能否稳定接管原本依赖人类团队持续协作的生产责任单元。

8. 结论

本文提出以复杂项目闭环能力作为 AGI 的工程化判断标准：在目标、资源、安全边界与初始上下文给定，且验收边界能够独立建立的前提下，AI 系统能够在无持续人工驱动的情况下，稳定完成多数原本需要普通人类团队长期协作的典型复杂项目，并提供可验证的交付结果。

该定义不试图取代认知型、形式化或社会影响型 AGI 定义，而是补充它们尚未充分覆盖的项目责任层。它把判断对象从孤立模型扩展到完整系统，把进展单位从局部任务扩展到复杂项目，并通过明确项目范围、人类基线、人工介入、资源条件和验收方式，使这一判断框架能够进一步操作化和比较。

当 AI 能够把广泛的局部能力持续组织为可靠、自治、可维护的复杂项目交付时，它才从高能力工具转变为真正的项目责任系统。以这一转变为判断单位，工程化 AGI 可以成为可比较、可验证并可逐步逼近的系统级工程目标。同时，将这一目标限定在可由外部条件闭合的复杂项目上，可以避免把项目执行自治与开放前沿创新混为一谈，并为进一步讨论 AGI 与 ASI 的工程化边界留下清晰接口。

参考文献

- [1] TURING A M. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460. DOI: [10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433).
- [2] SEARLE J R. Minds, brains, and programs[J]. Behavioral and Brain Sciences, 1980, 3(3): 417-457. DOI: [10.1017/S0140525X00005756](https://doi.org/10.1017/S0140525X00005756).
- [3] LEGG S, HUTTER M. Universal intelligence: A definition of machine intelligence[J]. Minds and Machines, 2007, 17(4): 391-444. DOI: [10.1007/s11023-007-9079-x](https://doi.org/10.1007/s11023-007-9079-x).
- [4] GRACE K, SALVATIER J, DAFOE A, et al. When will AI exceed human performance? Evidence from AI experts[J]. Journal of Artificial Intelligence Research, 2018, 62: 729-754. DOI: [10.1613/jair.1.11222](https://doi.org/10.1613/jair.1.11222).
- [5] OPENAI. OpenAI Charter[EB/OL]. 2018. <https://openai.com/charter/>.
- [6] MORRIS M R, SOHL-DICKSTEIN J, FIEDEL N, et al. Position: Levels of AGI for operationalizing progress on the path to AGI[C]//Proceedings of the 41st International Conference on Machine Learning. PMLR, 2024, 235: 36308-36321. arXiv: [2311.02462](https://arxiv.org/abs/2311.02462).
- [7] HENDRYCKS D, SONG D, SZEGEDY C, et al. A definition of AGI[EB/OL]. arXiv, 2025. arXiv: [2510.18212](https://arxiv.org/abs/2510.18212).
- [8] BURNELL R, YAMAMORI Y, FIRAT O, et al. Measuring progress toward AGI: A cognitive framework[EB/OL]. arXiv, 2026. arXiv: [2605.28405](https://arxiv.org/abs/2605.28405).
- [9] KARNOFSKY H. Some background on our views regarding advanced artificial intelligence[EB/OL]. Open Philanthropy, 2016. <https://coefficientgiving.org/research/some-background-on-our-views-regarding-advanced-artificial-intelligence/>.
- [10] KWA T, WEST B, BECKER J, et al. Measuring AI ability to complete long tasks[EB/OL]. arXiv, 2025, revised 2026. arXiv: [2503.14499](https://arxiv.org/abs/2503.14499).
- [11] STARACE G, JAFFE O, SHERBURN D, et al. PaperBench: Evaluating AI's ability to replicate AI research[EB/OL]. arXiv, 2025. arXiv: [2504.01848](https://arxiv.org/abs/2504.01848).
- [12] HUGHES E, DENNIS M, PARKER-HOLDER J, et al. Position: Open-endedness is essential for artificial superhuman intelligence[C]//Proceedings of the 41st International Conference on Machine Learning. PMLR, 2024, 235: 20597-20616. arXiv: [2406.04268](https://arxiv.org/abs/2406.04268).