

KSC2：ASI 的渐进结构框架——结构开放性、结构自主性与 AGI-ASI 的结构性差异

KSC2: A Progressive Structural Framework for ASI — Structural Openness, Structural Autonomy, and the Structural Difference between AGI and ASI

刘明

版本 1.2 · 2026 年 8 月

摘要

人工通用智能（Artificial General Intelligence, AGI）与人工超级智能（Artificial Superintelligence, ASI）通常主要通过能力覆盖范围和相对于人类的表现水平加以描述，但这种描述难以直接呈现：当既有问题结构不足以支持求解时，AI 能否自主形成更有效的问题表征、抽象方式和探索方向。本文关注一个进一步的问题：**在达到完整 ASI 之前，能否从任务结构层面对这一能力的进展进行持续观察和比较？**

本文从知识—结构—约束框架（Knowledge-Structure-Constraint Framework, KSC）出发，引入**结构开放性（Structural Openness）**，描述在固定任务协议下，完成任务所需的高层问题表征、抽象方式、关键变量、关系结构和探索方向中，有多少未在任务开始前由外部条件预先给出，而必须在求解过程中形成。结构开放性构成任务侧的结构需求坐标，而不是求解者能力。

本文进一步以**结构自主性（Structural Autonomy）**描述 AI 在缺少任务执行阶段的关键人类 S 输入时，自主形成、评价和修正完成任务所需高层结构的能力。结构自主性构成系统侧的结构能力坐标。据此，ASI 的结构性进展可以表现为：在保持结构生成有效性的前提下，AI 稳定进入结构开放性越来越高的任务，对关键人类 S 输入的依赖持续下降，并将这种能力扩展到越来越广泛的领域。

这一框架不替代经典 ASI 能力水平定义，而是描述完整 ASI 所需的一项重要能力维度，也不能单独衡量系统距离完整 ASI 还有多远。本文进一步提出，AGI 与 ASI 的差异不宜理解为“AGI 没有结构创新而 ASI 才有”，而应理解为两者对结构自主性的**覆盖范围与能力上界要求不同**：AGI 的核心目标是广泛自主覆盖普通人类通过长期协作能够完成的智力工作；完整 ASI 则还要求在广泛的高结构开放性任务中，自主进入过去主要依赖最优秀人类进行问题重构、抽象发现和原创结构生成的能力区域，并达到或超过人类最高水平。二者因此可以作为两条同步推进、相互作用但并非严格前后继的能力前沿。

关键词：KSC；人工通用智能；人工超级智能；结构开放性；结构自主性；结构生成；开放式智能；人机协同

一、引言

生成式 AI 正在快速扩展知识调用、推理、规划、代码生成和复杂任务执行能力。随着模型与智能体（Agent）系统的发展，过去需要人类大量参与的具体生成工作正在被逐步机器化，但人的重要作用并没有简单消失，而是越来越集中到少数决定问题理解和任务方向的高层节点。

KSC 框架将复杂目标任务中的智能活动理解为知识基础 K、结构生成 S 与约束保持 C 三个递进的功能阶次：K 提供任务可用的信息与状态；S 基于这些信息形成问题表征、关系、方案和行动路径；C 使既定目标、前提、边界和验收标准持续治理生成及状态转换 [1]。

在目标和验收条件相对明确的任务中，AI 可以大量承担 K 与 S，人的主要作用往往集中于约束保持。但当问题本身应该如何理解、哪些变量重要、什么抽象有效以及下一步应探索什么方向尚未确定时，人的高层指导已经不只是约束已有生成。重新定义问题、发现新关系或提出新的探索方向，本身就在产生新的 S。

这一差异尤其关系到如何理解 AGI 与 ASI 之间的能力推进。

AGI 至今不存在唯一统一定义。Morris 等按照表现水平与通用性建立 AGI 分级体系，使 AGI 可以被描述为持续进展而非单一二元节点 [2]；OpenAI Charter 则曾以“高度自主，并在大多数具有经济价值的工作上超过人类”描述 AGI [3]。Burnell 等进一步从认知能力分解出发提出 AGI 进度测量框架 [7]。

经典 ASI 定义更强调相对于人类能力上界的优势。Bostrom 将 superintelligence 描述为一种在几乎所有重要领域都显著强于最优秀人类大脑的智能，并明确包括科学创造力、一般智慧和社会能力 [4,5]。Hughes 等进一步提出，open-endedness——相对于观察者持续产生新颖且具有学习价值的成果——是人工超级智能的重要条件 [6]。

这些研究主要回答系统达到何种能力水平可以被视为 AGI 或 ASI，以及如何描述通往这些能力状态的总体进展。本文关注一个更具体的问题：

当既有问题结构不足以支持求解时，AI 能否自主发现、形成并验证更有效的高层结构？这种能力是否可以成为观察 ASI 逼近过程的一条独立维度？

二、结构开放性：当问题空间本身成为求解对象

2.1 复杂不等于开放

复杂任务并不一定具有高结构开放性。

一个大型软件系统可能包含数百万行代码、大量模块和长期协作，但许多局部任务拥有明确目标、成熟技术体系、可观察状态和可测试结果。任务规模虽然巨大，求解过程仍可以被拆分并逐步限定到相对稳定的问题空间。

相反，一个表述很短的问题也可能具有很高的结构开放性。例如：

- 一个异常现象究竟应该如何理解；
- 一项长期停滞的研究是否从一开始就问错了问题；
- 两类独立现象是否具有共同机制；
- 当前理论使用的变量是否切分错误；
- 一个新产品真正应该解决什么问题。

这些问题的主要困难并非步骤太多，而是完成任务所需要的高层结构尚未被预先给定。

2.2 结构开放性取决于任务协议

本文将 结构开放性 (Structural Openness) 定义为：

在给定初始问题表征、外部提供的知识与工具、反馈条件和评价协议下，完成一个任务所需的有效问题表征、抽象方式、关键变量、关系结构和探索方向中，有多少未在任务开始前被预先指定，而需要在求解过程中形成。

这一关系可以概念性表示为：

$$O(T \mid R_0, K_0, E)$$

其中：

- T 表示目标任务；
- R_0 表示任务开始时外部给定的问题表征和结构；
- K_0 表示任务协议中由外部提供的知识、工具与初始状态；
- E 表示反馈、验证和评价环境；
- O 表示在这些条件下仍未预先给出的高层结构需求。

这里的 K_0 不表示被测系统参数中已经内化的全部训练知识。模型内部积累属于系统能力的一部分。如果某种高层结构没有在当前任务协议中被外部给出，而系统能够依靠自身能力检索、恢复、重构或重新生成该结构，这种差异应反映在系统的结构自主性上，而不是反过来改变任务本身的结构开放性。

因此，在比较不同模型或系统时，应尽量固定 T 、 R_0 、 K_0 与 E 。不同系统面对的是同一组任务条件，区别在于它们能否自主形成尚未给出的结构。

结构开放性由此不是脱离具体求解条件的绝对属性，也不是求解者能力的函数。它描述的是：在一个明确的任务协议下，还有多少关键高层结构需要在求解过程中形成。

它也不等同于任务是否存在唯一正确答案，更不等同于任务价值、规模或一般意义上的难度。

2.3 结构开放性目前更适合分级比较

本文目前不主张已经可以稳定计算 O 的统一数值。

未被预先给出的结构需求可能同时涉及多个方面：

- 是否已经给出正确的问题表征；
- 是否已经给出重要变量；
- 是否已经给出候选关系；
- 是否需要从多个候选结构中选择；
- 是否需要主动推翻已有结构；
- 是否需要改变抽象层级；
- 是否需要提出任务协议中原本不存在的探索方向。

这些变化并不天然能够压缩到同一个标量。因此，现阶段更适合在控制任务条件的情况下，对结构开放性进行分级比较，并形成局部可比较关系。

例如，对于同一个底层问题，可以构造：

直接给出正确问题表征

→ 给出多个候选表征，但不指出正确结构

→ 只给出现象、材料与目标

→ 初始问题描述本身包含需要主动推翻的误导结构。

在这种受控任务族中，后续版本要求系统承担更多的问题形成、结构选择与结构修正责任，可以认为其结构开放性要求提高。

因此，本文使用“更高结构开放性”时，首先指这种受控条件下结构需求的增加，而不意味着已经存在跨所有任务统一可计算的尺度。

2.4 从搜索答案到构造问题空间

在结构开放性较低的任务中，系统通常已经拥有相对稳定的问题表征、成熟的方法体系、明确的目标边界以及能够淘汰错误路径的反馈机制。

此时，求解的主要问题是：

在给定问题空间中，什么路径可以得到有效答案？

随着结构开放性提高，当前问题表示本身可能错误，重要变量尚未被发现，候选关系需要重新形成，甚至评价维度也可能在探索过程中改变。

此时问题进一步变成：

应该建立什么样的问题空间来寻找答案？

后一问题本身属于结构生成 S。

因此，结构开放性描述的是任务从“主要在既有问题空间中求解”逐步进入“问题空间本身也需要在求解中形成或修正”的程度，而不是对工程、科研或创造活动进行类别划分。

三、从约束协同到结构协同

3.1 人类的高层指导并不都属于 C

在 KSC 中，约束保持 C 的核心是使已经形成的目标、前提、边界和验收标准持续治理后续生成 [1]。

例如：

当前方案违反已经确定的性能要求。

主要属于 C。

当前结果没有通过既定测试，需要回退。

同样主要属于 C。

但如果人提出：

这可能并不是性能问题，而是状态一致性问题。

系统获得了新的问题表征。

如果提出：

两种过去分别处理的现象可能具有同一个底层机制。

系统获得了新的关系结构。

如果判断：

继续沿当前方向优化没有意义，需要从更高抽象层重新定义问题。

系统的后续搜索空间本身发生了改变。

按照 KSC 的功能边界，这些行为直接参与 S，而不应统一归入 C。

3.2 人机 S 协同

当前 AI 的突出能力之一，是能够以较低边际成本大规模展开一个已有结构。

一旦获得有价值的问题表征、假设或探索方向，模型可以快速完成知识关联、推理展开、反例寻找、结构比较、局部验证和实现推导。因此，人类不再需要亲自完成一次关键抽象之后的全部认知工作。

开放问题中的一种典型人机协同可以表示为：

关键结构生成 → AI 大规模展开 → 新知识与候选结构形成 → 更高层结构生成 → 再次展开。

这里的关键结构既可能来自人，也可能来自 AI。

在许多当前系统中，真正稀缺的可能不是推理总量，而是少数能够显著改变后续问题空间的高价值 S 。近期自主科学研究也表明，自动化执行科研流程与自主形成、挑战和修正决定研究方向的高层结构，仍是不同层次的能力 [8-10]。

人机协同由此出现一个重要变化：

从人主要约束机器在既有结构中生成，进一步发展为人与机器共同形成新的问题结构。

四、结构自主性与 ASI 的结构进展

4.1 结构自主性

本文用 **结构自主性 (Structural Autonomy)** 描述系统在面对未预先给出的高层结构时的能力：

结构自主性，是系统在缺少任务执行阶段的关键人类 S 输入时，自主形成、评价和修正完成任务所需高层结构的能力。

这里的“形成”可以包括从已有知识中恢复、重构或组合有效结构，也可以包括产生新的问题表征、变量、关系和解释框架。

结构自主性不同于智能体的自主运行时间。

一个系统可以连续自主工作数天，却始终停留在错误的问题表示中；另一个系统则可能通过一次很短的推理发现原有问题定义错误，并形成更有效的抽象。

因此，结构自主性关心的不是 AI 能够自己执行多久，而是：

当现有结构不足以解决问题时，AI 能否自己形成更有效的结构，并以外部证据验证这种改变？

本文只考察系统在当前任务执行过程中的结构形成能力，并不要求这些结构在人类知识史上首次出现。

如果某个结构在人类知识史上已经存在，甚至可能出现在模型训练材料中，但没有作为当前任务协议的一部分直接提供给系统，而系统能够自主恢复、组合或重构它并据此改变后续问题空间，这种能力仍然可以计入结构自主性。

如果要证明系统产生了人类历史上此前不存在的原创结构，则需要更严格的新颖性和知识来源验证，不属于本文定义的必要条件。

4.2 关键人类 S 输入

本文所讨论的结构自主性并不意味着 AI 可以脱离人类文明积累独立产生全部认知结构。

模型训练数据、工具设计、语言、科学理论、任务定义和评价协议本身都包含大量历史人类结构。因此，本文观察的是一个更有限、也更可操作的问题：

在给定模型、工具、初始任务材料和评价环境后，完成当前任务还依赖哪些关键人类 S 输入？

这里的“关键”具有反事实意义，而不能只按提示次数机械计数。一次问题重构可能比数十次局部提示更重要；多个单独并不决定成败的提示，也可能共同改变任务结果。

因此，可以通过移除、替换或延迟某个人类结构输入，观察任务成功率是否显著下降、系统是否持续停留在错误问题表示中，或原本可达的有效求解路径是否不再出现。

同时需要区分验证与新的结构输入。

如果人类仅依据预先确定的评价协议返回“通过 / 不通过”或外部实验结果，这主要是在提供验证或维持约束。但如果验证者进一步告诉系统问题表示为什么错误、应该引入什么变量、改变什么抽象层级或探索什么方向，验证已经同时向系统提供了新的 S。

因此，“AI 提出结构、人类只验证”只有在验证过程不额外提供关键结构方向时，才能被视为较低的人类 S 依赖状态。

4.3 如何观察 ASI 的结构性进展

经典 ASI 定义主要回答：

什么样的系统可以被称为超级智能？

本文补充的问题是：

在尚未达到这一终点时，什么变化可以被视为沿着 ASI 所要求的高阶结构能力方向前进？

本文提出：

ASI 的结构性进展，可以表现为 AI 在越来越广泛、结构开放性越来越高的任务中，对关键人类 S 输入的依赖持续下降，并逐步获得自主形成、评价和修正高价值结构的能力。

结构自主性并不等同于完整 ASI。一个系统即使在狭窄领域具有很高结构自主性，也不能因此被称为完整 ASI；经典 ASI 定义仍要求系统在广泛重要领域达到极高能力水平。

本文因此只主张：如果一个系统最终要在广泛重要智力领域达到或超过人类最高水平，那么它很可能需要越来越少依赖人类为具体任务提供决定性的问题重构、抽象发现和高层方向选择。

4.4 四个观测维度

本文现阶段不把 ASI 的结构性进展压缩为单一数值，而是提出四个可以逐步实验化的观测维度。

第一，任务的结构开放性

在固定或记录 R_0 、 K_0 与 E 的情况下，观察 AI 能够稳定处理结构开放性达到何种程度的任务。

第二，对关键人类 S 输入的依赖

对于同一任务，可以观察系统从：

人产生 S，AI 展开

逐步进入：

AI 提出候选 S，人选择

→ AI 提出和筛选 S，人负责验证

→ AI 自主提出、筛选、验证并替换 S。

这一变化可以直接反映结构自主性的提高。

第三，结构生成的有效性

产生更多方向本身不能证明结构自主性提高。

有效结构需要通过外部证据体现价值，例如解释更多事实、显著压缩问题、产生新的可验证预测、打开此前不存在的有效搜索方向，或提高实际工程、科研与决策结果。

第四，能力覆盖范围

局部领域出现超人结构生成并不等于完整 ASI。还需要观察这一能力能够覆盖多少不同任务和领域，以及能否逐步扩展到广泛人类智力活动。

这四个维度可以共同描述 ASI 结构能力的进展，而无需立即合成为单一分数。

五、与既有 AGI、ASI 和开放式智能研究的关系

5.1 AGI 的渐进描述

Morris 等与 Burnell 等已经分别从能力水平、通用性和认知能力画像讨论 AGI 的渐进测量 [2,7]。因此，本文并不把“AGI 可以渐进描述”本身作为新的贡献。

本文进一步讨论的是：

能否为 ASI 的逼近过程增加一个与任务结构和结构自主性有关的连续坐标。

5.2 经典 ASI 定义强调最终能力水平

Bostrom 的经典 superintelligence 定义把最优秀人类作为比较基准，并要求系统在几乎全部重要智力领域达到显著更高水平，其中明确包括科学创造力、一般智慧和社会能力 [4,5]。

因此，本文并不主张既有 ASI 研究忽略创造力、科学发现或高阶认知。

本文提出的不同问题是：

这些高阶能力为什么会对问题结构提出不同要求，以及这一要求能否在真正达到 ASI 终点之前被连续观察？

许多复杂工作可以在成熟的人类认知结构内完成；而人类最高层次的科学、理论与创造活动中，一部分关键推进发生在旧结构本身已经不足的时候。

从 KSC 视角看，这意味着 AI 不仅需要在给定问题结构中表现得更好，还需要提高结构本身成为求解对象时的自主能力。

5.3 结构开放性不等于 open-endedness

Hughes 等提出，open-endedness 是人工超级智能的重要属性，并把它形式化为相对于观察者持续产生新颖且具有学习价值的成果 [6]。

本文的结构开放性与这一概念邻近，但描述对象不同。

open-endedness 主要关注：

一个系统或过程能否持续产生新的、对观察者而言具有学习价值的成果。

结构开放性主要关注：

完成一个具体任务时，有多少关键问题结构没有被任务协议预先给出，而必须在求解过程中形成。

一个开放式系统可以持续产生新成果，但其中某一次具体任务仍可能拥有较低结构开放性；一个高结构开放性的任务也不一定要求系统长期无界地产生新颖成果。

结构自主性则进一步关注：

当这些关键结构没有被预先给出时，AI 能否在缺少关键人类 S 输入的情况下自主形成、比较、验证和替换它们？

因此，本文并不主张首次把“开放性”与 ASI 联系起来，而是补充一个面向具体任务求解的结构坐标。

5.4 固定表示框架与结构开放性

Cao 与 Yang 指出，当前 AI 的推理、搜索与验证通常发生在预先给定的表示框架中，并区分了在既有表示空间内进行的变换与能够改变表示空间本身的生成性变换；他们据此提出**表征词汇缺口**

(**vocabulary gap**) 与**验证器缺口**(**verifier gap**)，分别描述新表示原语的生成与稳定，以及在其长期价值尚未充分显现时进行评价的困难 [11]。

这一问题与本文高度邻近：二者都强调，高阶开放式智能不能只在既有问题空间内搜索，还需要在既有表示不足时改变问题空间本身。但两者描述的坐标不同。Cao 与 Yang 主要讨论系统能否扩展自身的表示框架及相应验证机制；本文进一步将任务侧的**结构开放性**与系统侧的**结构自主性**分离：前者描述固定

任务协议下还有多少关键高层结构未被预先给出，后者描述 AI 能否在缺少关键人类 S 输入时自主形成、评价和修正这些结构。

因此，本文不把“超越固定表示”本身视为独有主张，而是进一步把这一问题转化为可在受控任务协议中比较的任务结构与系统能力变量，并用于描述 ASI 的渐进结构进展。

5.5 自主科研与结构自主性

近年的自主科学研究正在直接进入与结构自主性相邻的问题区域。

Virtuous Machines 展示了通用智能体系统贯穿研究工作流的能力，同时指出人类仍在创造性问题形成、概念创新等环节提供重要价值 [8]。

InquiTree 将科学探究过程建模为包含假设形成、研究设计、结果解释和信念更新的交互过程，并观察到智能体在长期科学探究中的判断与异常识别仍存在明显限制 [9]。

Socratic Agents 相关工作则使用**认识自主性 (epistemic autonomy)** 描述系统根据证据自主构造、挑战和修正物理解释的能力，尝试把自主科研从执行固定流程推进到解释框架本身的自我修正 [10]。

这些工作与本文并不采用同一理论框架，但共同指向一个重要区别：

自动化执行科学流程，与自主形成和修正决定研究方向的高层结构，并不是同一层能力。

本文的结构自主性试图把这一问题从自主科研扩展为一个更一般的跨任务结构能力变量。

六、AGI 与 ASI 在结构自主性要求上的差异

6.1 两种目标能力

在本文的任务结构语境中，**人工通用智能 (AGI)** 的目标可以描述为：

AI 能够自主承担广泛的、通常需要普通人类通过长期协作才能完成的智力工作，并形成持续有效的任务闭环。

这一表述用于讨论本文的任务结构问题，并不试图取代现有全部 AGI 定义。

它关注的不是单项 benchmark，而是 AI 能否自主承担过去需要普通人组成团队，通过持续规划、执行、检查和修正才能完成的工作。

相应地，从任务结构角度，**人工超级智能 (ASI)** 的目标可以进一步描述为：

AI 能够自主承担几乎全部人类能够完成的重要智力工作，并在广泛的、高度依赖问题重构、抽象发现和原创结构生成的任务中达到或超过最优秀的人类。

因此：

AGI 的核心目标是普通人类长期协作能力的广泛自主覆盖；完整 ASI 则进一步要求这种覆盖延伸到人类最高水平的结构发现、重构与原创生成活动。

6.2 差异在于结构自主性的覆盖范围与能力上界

本文不主张 AGI 完全没有创新能力，也不主张只有 ASI 才能产生新结构。

任何足够通用的 AGI 都可能在部分任务中完成新的问题表示、抽象和结构生成；当前系统也可能在若干窄域显示出不同程度的此类能力。

因此，AGI 与 ASI 的差异不应写成：

AGI 使用已有结构，ASI 才创造新结构。

更准确的表述是：

AGI 与 ASI 对结构自主性的最低覆盖范围和能力上界要求不同。

大量普通人类长期协作所完成的工作虽然十分复杂，却建立在人类文明已经形成的认知结构之上。成熟学科、工程体系、组织制度和专业实践已经提供问题分类、基础概念、方法体系、工具、工作流程、评价标准和历史经验。普通工作者无需重新发明这些结构，就可以通过学习与协作完成复杂任务。

因此，AGI 的广泛自主覆盖可以首先表现为可靠利用和组合既有人类结构，把复杂协作过程转化为机器可持续运行的任务闭环。

但完整 ASI 的经典目标要求覆盖和超越最优秀人类的高阶智力活动。人类许多最高价值的科学与创造性成果恰恰发生在旧结构不再足够的时候：

- 重新定义一个长期存在的问题；
- 引入此前不存在的重要变量；
- 把两个独立现象统一为一个结构；
- 建立新的理论；
- 发明新的技术范式；
- 找到此前无人意识到值得搜索的方向。

因此，完整 ASI 无法只通过“在相同任务结构上持续提高性能”得到充分描述。

从 KSC 视角看：

随着能力目标从广泛人类水平推进到全面超越人类最高智力水平，结构自主性从一种可以局部出现的能力，逐渐变成必须在更高结构开放性和更广领域中稳定成立的核心要求。

本文所谓 AGI 与 ASI 的“结构性差异”，因此不是两种智能机制之间绝对的类别二分，而是它们对结构自主性的覆盖范围与能力上界提出了不同要求。

七、AGI 与 ASI：两条同步推进的能力前沿

7.1 不存在严格的先后逻辑

AGI 与 ASI 存在不同的覆盖目标，并不意味着完整 AGI 必须首先实现，然后 ASI 才能开始发展。

现实系统可能长期处于交叉状态。

一个系统可能尚不能长期自主承担普通组织的大量日常工作，却已经在某类数学、程序设计或科学问题上表现出很强的结构发现能力。

另一个系统则可能能够高度自主地完成大量成熟工程任务，却仍然依赖人类决定真正值得研究的问题以及新的抽象方向。

因此，本文将二者理解为两条可以同步推进的能力前沿：

AGI 前沿主要观察 AI 能够可靠自主覆盖多少普通人类长期协作任务。

ASI 的结构前沿主要观察 AI 能够自主处理多高结构开放性的任务、其中多少关键 S 已经无需人类提供，以及这种能力能覆盖多广。

两者可以互相促进，却不存在严格的技术先后关系。

7.2 一个关于当前工程路径的可检验假设

大量软件工程、运营和常规知识工作虽然具有很高复杂度，但通常更容易获得明确目标、任务拆分、状态记录、自动验证和回退机制。

因此，即使任务本身十分复杂，其执行过程仍较容易在相对稳定的问题结构内被逐步治理。

高结构开放性任务则存在额外困难：如果真正的问题表示尚未形成，仅仅缩小已有搜索空间并不能解决问题；如果关键变量尚未发现，也无法预先定义完整验证条件。

据此，本文提出一个可检验假设：

在当前模型与工程基础设施条件下，AI 系统自主覆盖结构开放性较低或中等的复杂任务，可能快于稳定自主覆盖广泛的高结构开放性任务；后者将更直接受到结构自主性的限制。

这一假设可以通过受控任务协议、跨模型比较和关键 S 消融实验进一步检验。

它不构成 AGI 与 ASI 实现时间的预测，也不意味着局部领域的高结构自主性必然晚于广泛任务自动化出现。

八、研究方向与边界

本文的重点是提出一种观察 ASI 结构性进展的框架，而不是直接建立完整的测量体系。

8.1 建立结构开放性的受控任务协议

需要进一步描述和比较任务的结构开放性：哪些高层结构在任务开始前已经给定，哪些必须在探索中形成。

实验比较应尽量固定 T 、 R_0 、 K_0 与 E ，并将模型内部能力作为系统侧变量处理，避免因为被测系统不同而改变任务坐标本身。

一种直接方法，是构造同一底层问题的多级版本：

- 直接给出正确问题表征；
- 给出多个候选结构但不指出正确结构；
- 只给出现象与目标；
- 初始描述包含需要主动识别和推翻的误导结构。

这样可以首先建立结构开放性的受控分级，而不必立即构造统一标量。

8.2 测量关键人类 S 输入

对于同一任务和不同模型，可以记录：

任务成功仍然依赖哪些当前 AI 无法自主形成或恢复的关键人类 S 输入？

比单纯计数更严格的方法，是对这些输入进行消融、延迟或替换：

- 移除该提示后任务是否仍能成功；
- 延迟该提示后 AI 是否能自行产生等价结构；
- 给出错误高层结构后 AI 能否主动推翻；
- 一组单独非决定性的提示是否共同影响任务成功；
- 不同模型达到成功所需的人类结构输入是否变化。

如果模型升级后，同一个任务从需要多次人类重构，变成只需要人类选择候选结构，再进一步变成自主完成结构提出、筛选和验证，就形成一条直接的纵向能力证据。

8.3 验证 AI 自主结构的真实有效性

结构的新颖性或数量必须与外部验证结合，避免把随机生成更多候选方向误判为更高的结构自主性。

后续研究需要进一步回答：

- 什么证据足以说明新结构优于原结构；
- 如何区分真正的问题重构与语言层面的重新表述；
- 如何区分任务执行过程中的结构自主性与历史原创性；
- 如何评价跨长时间尺度才能验证的理论结构；
- 如何处理多个结构都暂时能够解释数据的情况。

8.4 观察能力的跨领域扩张

单个任务、单一领域乃至单种科学能力达到超人水平，都不能单独等同于完整 ASI。

需要进一步区分窄域结构自主性、跨任务可迁移结构自主性、跨领域结构自主性，以及广泛的高结构开放性任务中的稳定结构自主性。

完整 ASI 所要求的是这种能力向广泛重要智力领域持续扩展，而不是少量突出的局部案例。

8.5 研究边界

本文保留以下边界：

1. 结构开放性不是任务价值、规模或一般难度的同义词，而是由具体任务协议决定的结构需求。
2. 当前阶段更适合通过受控任务对结构开放性进行分级比较，而不是建立统一标量。
3. 工程任务不必然具有低结构开放性，科研任务也不必然具有高结构开放性。
4. 结构自主性关注当前任务执行过程中对关键人类 S 输入的依赖，不意味着 AI 脱离训练数据、工具和人类文明积累，也不等同于历史原创性。
5. 人类验证只有在额外提供关键结构方向时，才能与新的 S 输入区分。
6. 结构自主性是一条重要能力维度，但不是完整 ASI 的充分条件。
7. 本文不提出统一的 ASI 数值指标，也不试图完整衡量系统距离完整 ASI 还有多远；对 AGI 与 ASI 的描述也仅是任务结构维度上的补充。

九、结论

本文提出了一个用于描述 ASI 逼近过程的渐进结构框架。

其中，**结构开放性**描述固定任务协议下还有多少关键高层结构需要在求解过程中形成；**结构自主性**描述 AI 在缺少关键人类 S 输入时自主形成和修正这些结构的能力。前者构成任务侧的结构需求坐标，后者构成系统侧的结构能力坐标。

由此，ASI 的结构性进展可以被描述为：

AI 能够稳定进入结构开放性越来越高的任务，对关键人类 S 输入的依赖持续下降，同时保持结构生成有效性，并将这种能力扩展到越来越广泛的智力领域。

这一框架不是完整的 ASI 定义，也不能单独衡量系统距离完整 ASI 还有多远；它描述的是完整 ASI 所需的一项结构性能力维度。

它同时给出了本文对 AGI 与 ASI 结构性差异的回答：差异不在于 AGI 是否具有结构创新，而在于二者对结构自主性的覆盖范围和能力上界要求不同。AGI 的核心目标是广泛自主覆盖普通人类通过长期协作能够完成的智力工作；完整 ASI 则进一步要求这种能力进入并超越过去主要依赖人类最高水平问题重构、抽象发现和原创结构生成的区域。

因此，AGI 与 ASI 可以作为两条同步推进的能力前沿，而不是两个必须严格依次出现的离散事件。

本文最终把一个通常以 ASI 终点状态讨论的问题，转化为一个可以提前积累经验数据的问题：

在尚未被 AI 自主覆盖的任务中，哪些决定性的结构生成仍然必须来自人类？

如果这一问题能够通过固定任务协议、结构消融、跨模型比较与外部验证持续记录，就可以逐步观察 AI 从依赖任务执行阶段的人类结构输入，走向更高结构自主性的过程。

参考文献

- [1] 刘明. 知识、结构与约束：KSC 视角下的人类—AI 差异与协同. 2026. DOI: 10.5281/zenodo.21834579.
- [2] Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. **Position: Levels of AGI for Operationalizing Progress on the Path to AGI**. Proceedings of the 41st International Conference on Machine Learning, PMLR 235: 36308–36321, 2024. arXiv:2311.02462.
- [3] OpenAI. **OpenAI Charter**. 2018.
- [4] Bostrom, N. **How Long Before Superintelligence?** International Journal of Futures Studies, 2(1): 1–9, 1998.
- [5] Bostrom, N. **Ethical Issues in Advanced Artificial Intelligence**. In Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence, Vol. 2, pp. 12–17, 2003.
- [6] Hughes, E., Dennis, M. D., Parker-Holder, J., Behbahani, F., Mavalankar, A., Shi, Y., Schaul, T., & Rocktäschel, T. **Position: Open-Endedness is Essential for Artificial Superhuman Intelligence**. Proceedings of the 41st International Conference on Machine Learning, PMLR 235: 20597–20616, 2024.
- [7] Burnell, R., Yamamori, Y., Firat, O., Olszewska, K., Hughes-Fitt, S., Kelly, O., Galatzer-Levy, I. R., Morris, M. R., Dafoe, A., Snyder, A. M., Goodman, N. D., Botvinick, M., & Legg, S. **Measuring Progress Toward AGI: A Cognitive Framework**. arXiv:2605.28405, 2026.
- [8] Wehr, G., Rideaux, R., Fox, A. J., Lightfoot, D. R., Tangen, J., Mattingley, J. B., & Ehrhardt, S. E. **Virtuous Machines: Towards Artificial General Science**. arXiv:2508.13421, 2025.
- [9] Cui, S. **InquiTree: Evaluating AI Agents in the Scientific Inquiry Loop with Paper-Derived Research Trees**. arXiv:2606.09550, 2026.
- [10] Zeng, X., Liu, P., Zang, Y., Shen, Y., Yu, F., Yu, C., Liu, M., & Du, Y. **Socratic Agents for Autonomous Scientific Discovery in High-Dimensional Physical Systems**. arXiv:2606.26722, 2026.
- [11] Cao, Y., & Yang, H. **Beyond Fixed Representations: The Vocabulary and Verifier Gaps in Open-Ended AI**. arXiv:2607.09560, 2026.