

KSC：知识、结构与约束——人类—AI 差异与协同的功能框架

KSC: Knowledge, Structure, and Constraint —A Functional Framework for Human-AI Differences and Collaboration

刘明

版本 V3.3.1 · 2026 年 8 月

摘要

当前 AI 已能够广泛调用知识并快速生成复杂结构，但不同任务中的能力表现仍具有显著差异：有些问题主要受可用信息限制，有些问题要求形成新的问题表征、关系或方案，还有一些任务即使局部生成质量很高，也会在长链条、多条件和强状态依赖过程中逐渐丢失目标、前提、边界或验收标准。仅以单一的“智能强弱”描述这些差异，难以解释当前 AI 的能力边界及人机协同方式。

本文提出知识—结构—约束框架（Knowledge-Structure-Constraint Framework，简称 KSC），将复杂目标任务中的智能活动描述为三个递进的功能坐标：K 表示系统当前可访问或可调用的信息与状态；S 表示基于 K 形成问题表征、关系、方案和行动路径的结构生成；C 表示已有目标、前提、边界、不变量和验收标准对 S 及其引发的状态转换持续发挥约束作用。KSC 提供的是理解智能能力结构的低维坐标，而不是把具体行为或失败排他地归入 K、S 或 C 的诊断工具。

本文进一步分析 C 在有限的推理与生成过程中的作用。任何一次实际模型调用及其内部推理过程都有实际边界，复杂工作因而往往需要跨越多次调用、多个推理阶段和多次状态转换。单次调用内部的约束遵循与多次调用之间的一致性保持，可以理解为 C 在不同时间尺度上的表现：有限推理过程不仅形成新的结构，也需要使已有目标、前提和边界继续影响新的生成。约束信息进入 K 并不意味着 C 自动成立；只有这些条件在推理、行动和状态转换中实际发挥约束作用时，才构成 C。这里讨论的是功能关系，不预设独立的 C 模块、固定资源分区或特定搜索机制。

基于 KSC，本文提出一个阶段性比较判断：以当前大语言模型为核心的 AI 系统，在知识调用广度以及候选结构生成的速度、规模和迭代能力上具有明显优势；典型的人类工作系统则因长期处于现实环境、承担责任与行动后果，并借助文档、制度、测试、权限和状态记录等外部机制，在许多复杂任务中通常具有更稳定的约束保持条件。在目标、边界和验收标准相对明确的工程任务中，人的关键增量经常集中在 C；在问题表征和探索方向尚未形成的开放问题中，人类也可能直接参与 S。人类、AI 与外部系统可以共同承担 K、S、C，由此形成不同的人机协同结构。KSC 不涉及意识、主体性、价值起源或模型底层计算机制的完备解释。

关键词：KSC 框架；人机协同；大语言模型；知识基础；结构生成；约束保持

一、引言

生成式 AI 的使用正在从单轮问答和局部内容生成，扩展到软件开发、研究、规划和智能体等更长程的复杂工作。模型也越少作为孤立组件工作，而是与检索、工具、外部状态、验证机制和人类判断共同组成完整任务系统。由此，“模型一次能够生成什么”与“完整系统能够持续完成什么”之间的差异，正在成为理解 AI 能力的重要问题。

现有讨论通常分别从知识不足、推理能力、上下文利用、指令遵循、外部记忆、长程规划或人机互补等角度解释这些差异。这些研究对应不同技术机制，但缺少一套足够简洁的功能语言，把它们放在同一能力结构中讨论。结果是，“模型能力”或“智能水平”往往同时指向知识、生成、控制和系统协同等不同问题，难

以稳定比较不同模型、任务与人机系统。

因此，本文关注三个相互关联的问题：能否建立一套低维能力坐标，用统一语言描述复杂任务中不同类型的智能功能；当任务超出一次有限的推理与生成过程能够完成的范围时，已有目标、前提、边界和验收标准如何继续作用于后续过程；以及这些能力如何在人类、AI 与外部系统之间组合，从而形成不同的人机协同方式。

本文主要讨论目标和评价标准已经形成或经协作确认、旨在形成可交付结果或改变外部状态的复杂任务。

二、KSC：知识、结构与约束的三阶功能框架

2.1 总体结构

KSC 将复杂目标任务中的信息活动分为三个递进的功能阶次：

功能阶次	KSC 功能	核心问题	典型对象或表现
一阶：可用信息	K：知识基础 (Knowledge Base)	系统当前能够获得并调用什么？	参数知识、长期记忆、上下文、检索结果、现场事实、任务状态与经验记录
二阶：结构生成	S：结构生成 (Structure Generation)	系统如何基于信息形成新结构？	抽象、分类、关系发现、问题表征、推理、分解、计划、方案与修正路径
三阶：生成控制	C：约束保持 (Constraint Maintenance)	什么持续约束生成与状态转换？	目标、前提、边界、权限、不变量与验收标准对生成、行动和状态更新的持续控制

三者的基本关系是：K 为 S 提供材料；S 将可用信息组织为新的问题结构、关系和候选行动；C 依据已经形成的有效条件，对 S 及其引发的状态转换进行接受、限制、修正、回退、授权更新或验收。

这里的“三阶”表示功能依赖关系，而不是固定的时间顺序、三个彼此独立的物理模块或数学意义上的高阶变量。K、S、C 可以由同一次模型调用、同一个人类主体或同一个外部系统共同承担。一次评估结果可以成为后续可用的 K；依据既定条件进行判断时，同一推理过程可以承担 C；形成新的修正方案时又可以承担 S。

KSC 因而是一套能力坐标，而不是组件分类表。它关注信息和计算在当前任务中承担什么功能，以及完整系统在三个功能方向上具有什么能力条件。

2.2 K：知识基础

K 是在当前任务阶段已经进入系统可访问或可调用范围、原则上可供后续处理使用的信息与状态。它既包括已经固化在模型参数或人的长期记忆中的知识，也包括上下文、外部检索、任务文档、现场事实、历史决策和当前状态。

信息客观存在，不等于已经成为当前任务可用的 K。只有当系统能够在相关阶段访问或调用它，它才进入当前任务的知识基础。信息是否被正确识别、赋予适当权重并实际用于后续生成，不由 K 本身保证。因此，K 的意义不只在信息总量，还取决于相关信息能否进入当前任务的可用范围。

K 丰富也不自动意味着能够解决新问题。已有信息只是后续结构生成的材料；面对新的关系、问题表示或行动路径，系统仍需要 S。

2.3 S：结构生成

S 是基于 K 形成新结构的功能，包括问题表征、关系发现、抽象、分类、因果推理、任务分解、规划、方案生成、搜索方向形成以及根据反馈产生新的修正路径。

S 不只表示输出一个答案，也包括改变问题的表示方式、发现新的变量和关系、提出候选解释或技术路径，以及在反馈后重新组织已有信息。一个系统可以拥有大量 K，却因为缺少关键结构而无法推进；也可以通过新的 S 使原本分散的信息形成可行动的问题结构。

S 的结果可以成为后续任务中的新 K，也可以改变系统接下来能够看到的问题空间。结构生成本身并不保证这些新结构始终符合既定目标、前提和边界，这一持续约束功能由 C 描述。

2.4 C：约束保持

C，即约束保持，是指已经形成的目标、前提、边界、优先级、不变量、权限和验收标准，在连续推理、行动和状态转换中持续发挥约束作用，并只能在授权或明确确认下被更新或解除。

目标、边界和验收标准作为被记录、保存和调用的信息时属于 K；当这些信息实际影响当前的信息选择、候选结构、行动许可、结果判断、状态更新或完成判断时，则发挥 C 的功能。因此，C 不是一组独立于 K 的特殊信息，而是已有条件对生成及状态转换持续发挥约束作用的功能关系。

C 也不等于简单记住要求。约束已经写入提示、文档、记忆或状态，并不意味着它已经生效。系统仍需在相关阶段识别其有效性、作用范围、优先级和修改权限，并据此限制、筛选、修正、回退或终止不符合条件的生成路径。

S 与 C 的区分不是底层计算操作的排他分类，而是功能角色的区分：S 关注系统形成了什么新的结构；C 关注已经成立的条件是否继续决定哪些结构、行动和状态转换可以被接受。执行 C 的具体过程完全可以包含推理、比较、分类等同样可能出现在 S 中的计算，也可以同时产生新的 S。

判断当前结构是否违反既定条件、是否需要修正或回退，可以承担 C；形成具体修正方案和新的行动路径则属于 S。当既有条件不足以确定问题表征或探索方向，需要形成新的抽象、关系或候选方向时，仍属于 S。同一复杂行为可以同时承担 S 与 C，而不需要被排他地归入其中之一。

约束也不是永久静止的。新事实可以进入 K，新的 S 可以形成对任务的新理解；经授权或明确确认后，目标、优先级和边界可以更新。被识别并授权的改变属于 C 的正常更新，未经识别的遗忘、弱化、覆盖或越权修改则属于约束失效或漂移。

2.5 分析层级与边界

KSC 描述的是复杂任务中的功能关系，而不是排他的故障分类。同一行为、工具或失败可能同时涉及 K、S、C，也可能在不同阶段承担不同功能；框架不要求把每一个实际问题唯一映射到某个字母，也不以提供项目级故障诊断流程为目标。

KSC 的作用是在更高层建立能力坐标：系统能够获得怎样的信息基础，形成怎样的新结构，以及已有目标、前提和边界能够在多大范围内持续约束任务推进。这个坐标可以用于比较不同模型、人类主体和混合系统，也可以描述同一系统在不同任务结构下表现出的能力差异。

三、KSC 视角下的当前人类—AI 能力差异

3.1 系统级比较

KSC 比较的不是人脑与模型的底层同构能力，而是完整任务系统在 K、S、C 三个方向上分别能够达到怎样的水平，以及这些能力由哪些主体和机制共同提供。人类可以依靠长期记忆、身体环境、工具、制度和社会关系；AI 系统则可以依靠模型、上下文、检索、记忆、执行器和验证机制。

因此，底层实现不同并不妨碍功能比较。需要观察的是：系统能够在多大范围内获得相关信息，能够形

成怎样的新结构，又能在何种规模和持续时间的任务中保持既有目标和边界有效。

3.2 当前 AI 的 K 与 S 特征

当前大语言模型拥有广泛的参数化知识，也可以接入检索、工具和外部文档。它能够以较低成本快速完成分类、概括、比较、分解、规划、解释、写作和代码生成，大幅提高单个任务中知识调用和候选结构生成的速度、数量与迭代能力。

一项包含 758 名知识工作者的预注册实验显示，在模型能力边界内，GPT-4 辅助能够提高任务完成数量、速度和质量；但在超出模型能力边界的复杂管理任务中，AI 辅助也可能降低正确率 [1]。这一结果说明 AI 辅助的效果仍受任务结构和模型能力边界限制；从 KSC 视角看，可以将其理解为知识调用与候选结构生成方面的优势能够显著提高局部任务能力，但并不足以保证超出能力边界后的整体表现。

这里所说的 S 优势主要指当前 AI 在候选结构生成的速度、规模和迭代能力上的优势，并不意味着其在所有开放问题中都能够独立找到正确的问题表征或探索方向。当决定性结构尚未形成时，人类仍可能提供关键的 S。

3.3 典型人类工作系统与当前 AI 系统在 C 上的阶段性差异

典型的人类工作系统同样受到个体知识覆盖、结构生成速度和工作记忆的限制，但人通常长期处于连续现实环境中，受到责任关系、社会评价、风险和行动后果约束，并持续借助文档、制度、测试、权限和状态记录维持任务连续性。这些条件使典型人类工作系统在许多现实复杂任务中通常拥有更多让 C 持续有效的机制。

一个工程师未必能一次生成最优架构，但通常会持续知道不能破坏生产环境、不能忽略失败测试、不能随意删除用户数据。一个管理者未必能形式化全部变量，却会不断受到资源、时间、责任和现实反馈约束。很多时候，典型人类工作系统相对于独立模型调用的可靠性优势，并不来自更快的结构生成，而来自目标和边界与现实环境持续绑定。

这种比较并不是对人类工作系统永久优势的判断。工作记忆和认知控制本身同样存在容量与稳定性限制 [2, 3]；人类之所以能完成长程复杂工作，也大量依赖外部工具和组织机制。随着模型内部控制、持久状态、验证和外部治理机制发展，完整 AI 系统能够承担的 C 也会继续扩大。

3.4 任务结构决定人机优势如何组合

在目标、边界和验收标准相对明确的工程任务中，当前 AI 往往已经能够承担大量 K 与 S，因此人的关键增量更容易集中在 C，例如目标维持、边界控制、状态连续性和现实验收。

在更开放的问题中，问题表征、抽象方式和有效探索方向本身可能尚未确定，人类的作用会从 C 扩展到直接参与 S。由此，人机分工不是固定角色，而是不同任务结构下 K、S、C 能力条件的不同组合。

KSC 的价值就在于提供一个比“人思考、AI 执行”或“AI 生成、人监督”更一般的坐标，使这种阶段性差异能够被统一描述。

四、C 的进一步分析：有限推理过程与长程任务连续性

4.1 有限推理与生成过程

任何实际智能过程都受到时间、计算、信息访问和行动范围的限制。对当前大语言模型而言，一次调用能够容纳的上下文、完成的推理、生成的输出以及和外部环境交互的范围都有实际边界。即使这些边界持续扩大，只要单次过程仍然有限，就仍会存在超出一次局部生成能够完整完成的任务。

因此，复杂工作通常必须被展开为多个局部过程。系统可以多次调用模型，也可以在一次调用内部组织多个推理、验证和修正阶段，还可以在模型、工具和人类之间反复传递状态。每一个局部过程都可能表现出很强的知识调用和结构生成能力，但长程任务还要求此前已经形成的有效条件能够持续影响后续过程。

4.2 单次调用与跨调用中的 C

C 并不是跨调用之后才出现的能力。即使在一次较长的模型调用内部，前面已经给定或形成的目标、前提和边界也必须持续影响后续推理和生成。随着生成继续进行，模型一方面形成新的结构，另一方面还需要依据已有条件选择、判断、抑制或修正这些结构。

从 KSC 的功能角度看，S 与 C 可以由同一次有限推理过程共同实现。相关过程形成新的问题表示、关系、方案和行动路径时承担 S；使已有目标、前提和边界继续影响这些生成时承担 C。约束信息已经存在，并不意味着 C 自动成立；只有这些条件在实际推理过程中继续影响选择、判断、行动许可和状态转换时，才发挥 C 的功能。这里讨论的是功能角色，不意味着模型内部存在显式的 S/C 算力配额、可分离的资源分区、特定搜索过程、固定注意力分区或专门的 C 神经模块。

当任务跨越多次调用时，同一功能要求仍然存在。新的调用需要通过上下文、外部状态、记忆或其他机制，使此前仍然有效的条件重新成为当前任务可用的 K；这些条件还必须在新的推理过程中实际发挥约束作用，才能继续形成 C。

因此，单次调用内部的约束遵循与多次调用之间的一致性保持，可以理解为 C 在不同时间尺度上的表现。

4.3 长程任务中的连续性

当一个复杂任务经过多个局部阶段推进时，后续阶段面对的已经不再只是最初的任务描述，而是包含此前决策、状态变化、已排除路径和验收进展的持续任务状态。如果这些已经形成且仍然有效的条件不能继续影响新的生成，系统即使持续产生局部合理结果，也可能重新打开已经排除的路径、改变未经授权的边界、忽略已确认决策，或者把局部完成误判为整体完成。

因此，在长程任务中，C 描述共同目标、前提、边界和验收条件如何持续约束多个局部生成。它不能保证任务一定成功，系统仍需要足够的 K、S 和环境交互能力；但局部生成质量本身也不足以描述长程任务能否保持一致性。

4.4 C 视角下的若干失效层级

从 C 视角看，相关困难并非来自单一机制。即使目标和约束已经进入当前上下文，它们也未必稳定影响生成过程。长上下文中的信息位置效应 [6]、基于 Stroop 冲突范式观察到的控制性能退化 [7]，以及参数知识与上下文知识冲突 [8]，分别揭示了信息利用和控制中的不同限制；它们并不直接证明存在统一的 C 机制，但都与“相关条件已经可用却未稳定发挥约束作用”的问题相邻。复杂指令遵循基准显示，在包含多类约束、多任务以及复杂组合要求的设置中，当前模型仍存在明显的约束遵循困难 [13, 14]。

跨调用时，新的问题进一步出现。在不提供持久任务状态的独立调用中，一次调用中新形成的目标解释、决策、边界和验收状态不会自动成为下一次调用稳定可用的内部状态。长程对话和智能体系统因此常借助上下文重放、外部记忆或显式状态管理维持连续性；MemGPT 展示了通过分层记忆与虚拟上下文管理扩展跨会话连续性的一种实现 [9]。外部保存可以使相关信息继续存在，但这些信息重新进入上下文后仍需要被模型正确解释和遵循。

当任务经过更多状态转换、委派、修正和中间决策时，约束还可能在反复保存、继承、解释和执行中发生漂移。相关研究把约束在多智能体轨迹中逐渐丧失操作效力概括为约束漂移 [10]；MasDrift 进一步在受控多智能体任务中测量授权约束的继承，并观察到委派架构及层级深度与未授权行动、授权约束未被保持之间的系统性关系 [11]。上述概念主张与实验结果对应特定问题设置，不能直接外推为所有长任务的统一规律，但表明长程约束保持可以作为独立的能力问题加以研究。

五、KSC 视角下的人机协同

5.1 人、AI 与外部系统共同承担 K、S、C

传统的人机协同常被描述为“人提出需求，AI 执行任务”或“AI 生成，人负责监督”。这些表述能够描述某些表面分工，却难以覆盖不同任务中人和 AI 角色不断变化的情况。

从 KSC 视角看，复杂任务中的 K、S、C 可以由人类、AI 与外部系统共同承担。人类可以提供现场知识、目标和高层结构判断；AI 可以调用大规模知识并快速形成候选结构；外部系统可以保存事实和状态、执行形式化规则、提供测试与验证。不同主体承担什么功能取决于任务结构和当前能力，而不是预设的角色边界。

在本文重点讨论的工程任务中，一种常见情况是：AI 已经能够承担大量 K 与 S，人类和工程环境则承担更多 C。但当问题表征、技术方向或解决路径本身尚未形成时，人类也可以直接参与 S；当外部数据、状态和检索成为关键时，工具系统又会明显扩展 K。

5.2 外部系统如何扩展系统能力

外部工具可以同时改变系统的信息可用性、结构生成和约束保持能力。文档和检索可以扩大 K；计划工具、搜索和验证反馈可以影响 S；状态记录、权限、测试、审计和回退机制则可以让部分目标、边界和验收条件在多阶段任务中持续有效。

这些机制往往同时改变多个功能方向。更重要的是观察它们加入以后，系统在 K、S、C 三个方向上的能力如何变化。

对于 C 而言，外部化尤其具有直接意义。人类能够处理远超单次工作记忆范围的复杂项目，并不是因为所有目标、边界和状态始终保存在脑内。文档、制度、清单、版本记录、测试、权限、会议纪要和现实反馈长期承担了部分任务连续性功能。情境行动与延展认知研究也表明，人的认知和行动可以依赖环境与外部工具共同实现 [4, 5]。

AI 系统同样可以利用任务规格、状态记录、权限、状态机、测试、验证器、审计和版本控制，使部分 C 由持久的外部机制承担。其意义不只是增加模型能够看到的信息，而是让一些已经确认的目标、边界和验收条件能够在多个有限推理阶段之间继续发挥作用。

5.3 人机协同的结构性理解

从 KSC 视角看，评价 AI 的自主能力不能只观察模型在单次任务中的生成质量，还需要观察完整系统能否获得所需 K、形成足够 S，并通过模型或外部机制维持必要 C。模型、工具和工程机制的变化，都会改变系统在这三个方向上的能力。

KSC 因而提供的是理解人机能力如何组合和变化的坐标，而不是规定某种唯一的人机协作流程。

六、理论位置、边界与可检验命题

6.1 与认知理论和延展认知的关系

心理学期区分与既有知识相关的晶体能力和在新问题中发现关系、解决问题的流体能力 [12]；工作记忆研究关注信息的短期保持和操作 [2]；认知控制理论则强调主动维持目标及实现目标的方法，并据此指导后续思维和行动 [3]。

这些理论分别覆盖了 KSC 的重要局部功能，但 KSC 不把 K、S、C 等同于某三类心理能力。它分析的是完整任务系统中的功能关系：K 可以存在于模型参数、人的长期记忆、上下文或外部文档中；S 可以由人、模型或工具系统完成；C 也可以由认知控制、现实责任、权限、测试、状态机和治理流程共同承载。

情境行动与延展认知研究强调行动对环境和外部工具的依赖 [4, 5]。KSC 与这些研究相容，并进一步提供一套用于描述人、AI 与外部系统整体能力结构的功能坐标。

6.2 与大语言模型相邻研究的关系

长上下文利用、知识冲突、复杂指令遵循、外部记忆和约束漂移研究，分别揭示了当前 AI 在信息利用、指令保持和跨阶段状态继承等方面的不同限制 [6–11, 13, 14]。这些现象具有不同底层机制，不应被统一还原为上下文长度、记忆不足或单一控制缺陷。

KSC 提供的是更高层的共同描述语言：系统有什么信息可用，如何形成新的任务结构，以及已有条件如何继续影响任务推进。由此，可以在不假定底层机制相同的前提下，用同一功能坐标比较这些研究所揭示的能力差异。

6.3 KSC 的理论增量

KSC 的增量不在于首次提出知识、推理、认知控制或人机互补，而在于把可用信息、结构生成和约束保持组织为一套简洁的三阶功能坐标。

这一坐标使“AI 能力”不再只能作为单一尺度讨论。相同的总体表现可能来自不同的 K、S、C 组合；同一模型在不同外部系统和任务结构中，也可能表现出不同的完整能力。由此，知识调用、问题表征、结构生成、目标保持、状态连续性以及人机协作能够在同一理论语言下讨论，而不需要假定它们具有相同底层机制。

6.4 框架边界与经验命题

K、S、C 的三阶关系是本文提出的功能性划分。它的评价标准首先是定义是否自洽、功能关系是否清晰，以及能否在完整任务中提供有解释力的统一坐标；它不要求三种功能在统计上彼此独立，也不要求每次行为或失败被唯一归类。

由该框架提出的人类与 AI 比较、C 的具体失效机制以及外部系统的作用，则属于可以持续检验的经验性主张。后续研究可以围绕以下问题展开：

1. 不同任务类型中，人类与 AI 的 K、S、C 相对能力结构是否呈现稳定差异？
2. 当任务从短程局部生成扩展到长程复杂工作时，哪些能力条件开始成为主要限制？
3. 外部记忆、检索、规划、状态、权限、测试和验证机制分别如何改变完整系统的 KSC 能力表现？
4. 单次推理和跨调用过程中，目标、边界与已确认决策的持续有效性如何变化？
5. KSC 是否能够比单一的总体能力分数更好解释同一模型在不同任务结构和系统配置下的表现差异？

这些问题用于检验由框架导出的经验判断及其适用范围，而不是把 KSC 转化为项目级故障诊断方法。

七、结论

本文提出 KSC，将复杂目标任务中的智能活动描述为三个递进的功能坐标：K 表示系统当前可访问或可调用的信息和状态；S 表示系统基于这些信息形成新的问题表征、关系、方案和行动路径；C 表示已经形成的目标、前提、边界、不变量和验收条件如何持续约束新的生成及状态转换。由此，“智能能力”不再只有单一强弱尺度，而可以从信息基础、结构生成和约束保持三个方向理解。

在这一坐标下，当前人类与 AI 的差异可以得到更清晰的描述。以当前大语言模型为核心的 AI 系统，在知识调用广度以及候选结构生成的速度、规模和迭代能力上具有显著优势；典型的人类工作系统则因长期处于现实环境，并借助责任、反馈和大量外部机制，在许多复杂任务中通常具有更稳定的约束保持条件。在工程任务中，人的增量经常集中于 C；在开放问题中，人类也可能直接参与关键 S。这些都是阶段性的能力结构，而不是永久的人机分工。

本文进一步对 C 进行了重点分析。任何一次实际推理和生成过程都是有限的，复杂任务因此经常需要跨越多个局部过程。单次调用内部的约束遵循与多次调用之间的一致性保持，可以理解为 C 在不同时间尺度上的表现：已有目标、前提和边界必须在后续生成中继续发挥约束作用。模型内部控制、跨调用状态继承和长轨迹约束漂移展示了这一问题可能出现的不同层级。

C 也不必全部由模型内部承担。和人类长期依赖文档、制度、测试和现实反馈一样，AI 系统可以借

助外部状态、权限、验证、审计和回退机制扩展长程任务能力。更一般地说，人类、AI 与外部系统可以共同承担 K、S、C，并随着模型和工程条件变化不断重新组合。KSC 因而提供了一套理解智能能力结构的基础坐标，使 AI 的能力进展、能力限制以及人机协同方式能够在同一种功能语言中得到讨论。

参考文献

- [1] Dell’Acqua, F., McFowland, E. III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Kraye, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37(2), 403–423. <https://doi.org/10.1287/orsc.2025.21838>
- [2] Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- [3] Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- [4] Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press.
- [5] Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- [6] Liu, N. F., Lin, K., Hewitt, J., et al. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12, 157–173. https://doi.org/10.1162/tacl_a_00638
- [7] Patel, S. C., Wang, H., & Fan, J. (2026). Deficient executive control in transformer attention. *PNAS Nexus*, 5(6), pgag149. <https://doi.org/10.1093/pnasnexus/pgag149>
- [8] Xu, R., Qi, Z., Guo, Z., Wang, C., Wang, H., Zhang, Y., & Xu, W. (2024). Knowledge conflicts for LLMs: A survey. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 8541–8565. <https://doi.org/10.18653/v1/2024.emnlp-main.486>
- [9] Packer, C., Wooders, S., Lin, K., Fang, V., Patil, S. G., Stoica, I., & Gonzalez, J. E. (2023). MemGPT: Towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*.
- [10] Li, T., Ma, Y., Wen, H., et al. (2026). Safe multi-agent behavior must be maintained, not merely asserted: Constraint drift in LLM-based multi-agent systems. *arXiv preprint arXiv:2605.10481*.
- [11] Xu, Z., Zhang, X., Luo, H., Jin, Y., Dong, Y., & Salam, H. (2026). MasDrift: Benchmarking authorization preservation across multi-agent architectures. *arXiv preprint arXiv:2608.07556*.
- [12] Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253–270. <https://doi.org/10.1037/h0023816>
- [13] Zhang, T., Zhu, C., Shen, Y., et al. (2025). CFBench: A comprehensive constraints-following benchmark for LLMs. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 32926–32944. <https://doi.org/10.18653/v1/2025.acl-long.1581>
- [14] Zou, T., Zhang, X., Yu, H., Wang, M., Huang, F., & Li, Y. (2025). EIFBENCH: Extremely complex instruction following benchmark for large language models. *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 20930–20953. <https://doi.org/10.18653/v1/2025.emnlp-main.1059>