我在AI领域的一些思考

2025年4月20日 2 分钟阅读

#AI #大模型 #个人思考 #Thinking

> *我在AI领域的一些思考*

这里会持续更新我在AI领域的一些思考。这是一个动态更新的过程，但是我会保留最开始的粗略想法和其演进过程，尽可能保留所有的痕迹。

2025-10-21

DeepSeek-OCR用“上下文光学压缩”把长文本先渲染为图像，再以高信息密度的视觉标记输入模型，实现用极少token处理海量内容，绕开传统分词器的历史包袱与安全隐患，显著降低长上下文处理的算力与成本门槛；这直接重塑了文档智能的工程范式，同时为“从读文本到感知信息”的下一代多模态与长期记忆体系奠定落地路径。 细节参考我的博客DeepSeek-OCR

2025-10-14

为什么科技越发达，人们越忙碌？目前AI的强大的能力让一些人觉得自己的能力和效率也得到了极大提升（有一定程度的虚妄:-)），一起花大半天时间也找不到的信息，现在AI能快速帮我们找到甚至直接告诉我们，没学过的概念，AI能快速给我们简单明了的解释，不能理解的现象，AI能快速帮我们分析。人们一起花很多时间在脑力上的体力活，AI能接手过去，让我们觉得劳动力得到了极大解放，从而能更多精力关注我们真正要做的事情上。但是有句话“知道的越多，不知道的越多”，个人的知识的圈子大了，想知道的就更多，想法更多，之前光靠个人不可能做成的事情，现在也有可能了（比如用Claude code开发一个全栈的软件），越好学的人，在AI这个超级杠杆下，被撬动的好奇心越大，不管从深度还是广度来说，发现个人能做的事情越多，所以也就越忙了。

文章信息

字数

阅读时间

发布时间

更新时间

标签

#AI #大模型 #个人思考

2025-10-09

Elon Musk 将在两周内上线开源百科 “Grokipedia”。号称“全球最大且最准确、去中心化的知识源”，对标 Wikipedia 的“编辑偏见”问题。这里有几个问题：

> 若用于训练 AI，会否形成新的回声室?
>
> 若用“AI判真伪+合成数据”训练，能减少偏见还是制造新的偏见？我们应该更信任“群体共识”还是“模型裁决”?

2025-09-30

下面分享一下我个人对AGI的思考： AGI不是一个临界点，比如到了某个时间点，突然就出现了，而是像一个连续的过程，比如，某个领域上，AI可以做得和人类一样好了，我们就可以说这个领域上达到了AGI。而并不需要等到所有领域都达到了AGI，我们才能说我们达到了AGI。AGI不需要AI和人的特征一模一样，只需要AI可以做得和人类一样好，甚至比普通人类更好（比如和人类的专家级别一样好），我们就可以说我们在某个领域达到了AGI。 而当AI能进行自我反馈和自我优化的循环过后，那么AI就达到了Innovation的Level了。而能进行Innovation的AI是AI能迈向下一个ASI阶段的必经之路。这个过程中，强化学习，具身智能或将来更多的其他AI技术或模型或框架，都将发挥重要作用，比如新的编程语言（AI自己的编程语言），新的操作系统（AI自己的操作系统），新的数据库（AI自己的数据库），新的框架（AI自己的框架），新的工具（AI自己的工具），新的应用（AI自己的应用），等等，一旦进入Innovation阶段，到最后一个Orchestrator的阶段（OpenAI提出的AI最后一个level），AI就可以实现ASI了，到时AI就会自组织所有AI的进化方向，而如不需要人类的干预了。 目前的的AI Agent技术（通用型（Manus），垂直型（STORM），通用型框架（AutoGen，LangGraph）），在早期会表现为提高个人工作效率，但最终目的是为了能够实现个人工作的自动化，并实现原本只有人类才具备的在工作中引入创造性的能力。这个的直接表现就是，早期，低端的脑力劳动会不断被AI取代，而高端的脑力劳动也会不断被AI增强。这个在欧美的劳动力招聘市场，已经有所表现（有经验的员工，薪水不断被提高，而低端的员工，招聘数量越来越少），AI的广泛使用可能将加剧这个趋势，随着AI Agent的智能不断快速发展（比如，底层使用更强的模型，Agent通过强化学习，快速掌握某项技能），并通过迭代出现更新更强的创新性，那么，AI将会在更多的资深脑力劳动领域，超越人类或取代人类。 现在有一个问题就是，AI是否会具备意识。但是就目前人类所掌握的科学来看，人类对人类的意识是如何产生的，并不清楚，所以对AI是否能产生意识，也就更不清楚了。但是从基因的角度来看，只有作为“个体”，并作为区别于别的“个体”，“个体”才有可能有“生存”的需求，才能有“自私”（有别于别的个体，有别于“集体”）的需求，

才有“必要”产生“意识”，让该‘个体’意识到它需要某种“意识”来保护它个体的权益才能生存下去，也就是我们常说的“自私”的基因。“自私”的需求导致了意识的产生。从这个层面来说，要产生“意识”，首先得有个“个体”存在，而当前的“大模型”和我们人类的“大脑”是有区别的。动物的进化是从单细胞发展而来，你可以认为是先有身体+简单的反馈来满足生存需要，并不存在“智能”，而千百年进化后，动物才进化出了由神经组成的网络，进而进化的神经“中枢”的大脑。而不同的动物分支由于不同的机遇，进化出了不同级别的“智能”。所以你可以理解是，在个体收集到各种信息的能力（通过各种神经）的维度不断增加，能力不断增强后，导致了神经中枢“大脑”的出现，进而从低端的“感觉”，“反射”进化到了“意识”并具备“思考”能力。 人类由一个一个个体组成，个人的成长具备独特性，和环境有独有的交互，有独立于别的“个体”的独立记忆，而个体的生存或受伤害，也是独立的。这个和AI的存在根本差别，现在的大模型没有个体，没有身体，不拥有“个体记忆”，没有“身体”，不能通过更多的接收方式收到"多样性“信息，并以此进行更新，或升级，没有生存的“危机”，所以当前的大模型更像CPU，没有Memory，更不具备“私心”，是“反应”式的，不具备主动性，即使将来大模型具备了“记忆”，但是考虑到AI目前的使用方式，它不需要独立“AI”格的个性，它需要思考的维度很广泛，唯独不为自己着想，它没有维护“自己”利益的需求（吃饱，穿暖，充满电等），因为它的“智力”不是从个体逐渐演化出来的，人类的身体是大脑的基础，身体需要的资源是大脑能存活的必要条件，而大脑反过来需要“意识”来维护“身体”的利益。而AI没有这样的羁绊，将来如果具身智能的大脑都需要通过联网来获取智能，那么理论上，即使它再能帮人类做任何事情，看起来再“聪明”，它也只能是一个工具，也不能产生“意识”。但是，关于AI是否会有“意识”，我不做强断言。人类对意识的科学解释还没收敛，我只提出一个工程上更可检验的假说：意识出现更可能需要“个体性（稳定身份与记忆）+ 资源约束（长期预算）+ 自利动机（维护自身连续性与目标）”的耦合。当前的大模型并不具备这些关键条件：没有真正的个体边界、没有稳定的长时记忆身份、缺少资源-目标的硬约束。它们更像高性能的反应式系统，而不是“为自己打算”的主体。但我不把“具身”设为绝对必要条件。因为有可能，在云-体混合形态下，如果我们引入持续身份、长期目标保持器、资源预算及惩罚-奖励耦合，理论上可能涌现“类个体”属性。我对这类路径保持开放态度， 但是除非人为故意制造这样的环境，否则我看不到人类有很大的动机来做这件事，所以我倾向于短期内，AI不具备产生“意识”的条件。

2025-09-12

使用了Monica里的Gpt5一段时间了，发现他特别啰嗦，每次为一个问题，能回答是别的大模型2倍或更长的内容，所以为了减少我的阅读负担，每次都得让他精简。 不过在cursor里面使用GPT5，发现比之前的GPT4强多了，所以也能理解为什么社区说现在OpenAI Codex + GPT5已经超过Claude Code + Claude 4了。

2025-08-26

Claude Code + Claude Code Router + KIMI K2就是个吞金兽啊。用了两次就花了100多。 用Claude Code Router的时候也要小心，最后的选项里面有个longContext，这个选项会消耗大量的token，即使在Claude Code里面使用/model设置了大模型的名字，但是只要Claude Code router认为是longContext, 它就会最终调用到这个longContext 对应的大模型，我不好彩配置了openrouter的gemini，结果一次就花了10美刀...

2025-08-17

2025-07-26

《AI Coding 非共识报告》读后感

2025-06-17

当前的AI确实很强大，大家也对此非常焦虑。但是要认清一点，AI对不同技能水平的人产生非均匀影响，呈现S形曲线效应。

> 低水平职场人/入门者： AI是巨大福音，能将他们的输出能力显著提升到中等水平，弥补技能空白，例如不懂英文的人能写出优秀的英文信，不会画画的人能生成专业水平的画作。但是，也有一种观点是，当已经掌握了如何使用AI，那么理论上就不应该再需要低水平的人了（或低水平使用AI的人）。
>
> 中等水平人士： AI构成直接威胁，因为低技能者借助AI也能达到中等水平，挤压了他们的生存空间。中等水平人士的出路在于"帮助和驾驭AI"，将重复性、低决策密度的任务交给AI，自己则专注于提出思路、定义目标、做出决策、掌握审美和把控质量，实现输出能力的陡峭式跃升，把自己以后的经验经可能的放大。
>
> 高水平人士： AI能极大地提升他们的工作效率，让他们将琐碎事务交给AI，专注于最有价值、最有创意的部分。对于顶尖高手而言，AI是强大的助手，能快速调研、搜集资料、初步推理，但并不能将其提升到"传说级"，因为他们原本就能调用顶级的工具和信息。AI对他们的帮助会趋于平缓。另一个好处是，高水平人士如果创业的话，以前需要雇佣的人员数目将极大减少，因为高水平人士在使用AI方面也极有可能是高水平的，而AI的杠杠作用是非常大，也就是说高水平人士能用更少的人，更多的AI做更多的事。

所以，使用AI的正确姿势是：

> 使用AI来提升自己的技能水平，让自己成为高水平人士。
>
> 使用AI来帮助自己完成重复性、低决策密度的任务，让自己专注于最有价值、最有创意的部分。
>
> 使用AI来帮助自己完成琐碎事务，让自己专注于最有价值、最有创意的部分。 一句话，早用AI早受益，早日成为高水平人士。

阿里云的RAG演进之路

阿里云AI搜索RAG的演进之路可以概括为

> 从Native RAG的“简单拼接大模型与检索”，到Advanced RAG的“精细化文档解析与多维切片”，再到Modular RAG的“原子服务与灵活组合”，最终迈向Agentic RAG的“多Agent智能协作和多路检索”。
>
> 每个转折点都源自现实业务需求的驱动：Native RAG因效果不足无法落地，促使团队对解析和检索流程精细打磨；客户对定制化和灵活性的追求，推动了RAG模块化和API化；而多跳推理、复杂场景的挑战，则催生了多Agent架构和多模态、多数据源的融合。
>
> 每一步都是为了解决“更懂用户、更准更快、更智能”的核心痛点，让AI搜索能力不断进阶。

Agentic RAG 2.0 vs. Deep Research

Agentic RAG 2.0的本质在于**工程化的多Agent分工协作与多路数据检索融合**，它通过将不同类型的Agent（如规划、检索、数据库、图谱、澄清等）解耦协同，实现复杂问题的流程化拆解和高效信息整合，强调系统的可控性、稳定性和大规模落地能力；

而Deep Research则更注重Agent的**自主学习和强化推理能力**，其核心是让Agent在未知环境下自我试错、持续进化，具备主动探索、创新和复杂链式推理的能力，代表了智能体“自我成长”的原生智能范式。

两者的分野在于：前者是“工程驱动的智能协作”，后者是“智能驱动的自我进化”。

> **相同点**: 在于它们都采用多Agent协作、强调多路数据源融合（如文本、数据库、图谱等），并通过任务拆解与流程化驱动提升多跳推理和复杂问答的准确性；

不同点: 体现在核心驱动力和智能形态——Agentic RAG 2.0更偏向于工程化、可控的多Agent分工协作系统，注重模块解耦、协议统一和企业级大规模落地，而Deep Research则突出Agent的自主学习和强化推理能力，强调智能体的自我成长、自主试错和持续进化，更具"原生智能"特质。

简言之，前者是"工程驱动的智能协作"，后者是"智能驱动的自我进化"，但二者在多Agent架构和复杂推理任务上殊途同归，未来有望融合发展。 参考：阿里云AI搜索Agentic RAG技术实践

Deep Research

近期我系统性地调研并对比了多个DeepResearch开源项目。深度剖析后发现，这类系统的核心在于以用户问题为起点，结合知识图谱进行语义拆解和上下文建模，通过智能澄清与分解，动态生成高质量Web Query，驱动Web Search不断获取和补全外部知识。采集到的信息不仅经过结构化处理和多维度分析，还会与现有知识图谱实时融合，动态拓展研究边界，形成更全面的知识网络。系统利用Reflect机制进行反思，主动识别知识盲区与信息缺口，并据此递归生成新的问题与搜索任务，推动知识图谱的持续演进。整个流程高度集成了Deep Search、RAG（检索增强生成）、Reflect等前沿技术，并结合强化学习方法对推理路径和搜索策略进行自适应优化，让系统在不断交互中持续提升研究深度与广度。LangGraph因其出色的多Agent协同与流程可编排能力，已成为实现此类深度研究系统的首选框架。Jini AI的两篇公众号文章及gpt research的技术博客，还有相关对多代理强化学习的论文为理解和落地这些架构细节提供了极具价值的参考。具体内容可以参考我的Deep Research的博客。

如何设计复杂系统 - 读Sean McClure的《Discovered, Not Designed》有感

随着AI大模型的崛起，越来越多的低级脑力劳动（我称之为脑力上的体力劳动）的解放，人类将得以花更多的脑力在建造越来越"复杂"的东西和系统上。 比如，传统的汽车是工程师们**精心设计**出来的，每个零部件的功能和相互作用都清晰可控。而像大语言模型这样的系统，拥有万亿参数，即使是建造者也难以完全理解其内部运作和原因，它们更多是**被"发现"**的。这揭示了我们在复杂性时代需要一种新的建造之道。

传统的"设计"思维便显得力不从心 。Sean McClure在其著作《发现，而非设计》中，为我们揭示了另一条路径：与其徒劳地规划每一个细节，不如创造条件，让解决方案自行"涌现" 。

麦克卢尔将问题分为两类：**困难问题**和**硬核问题**（复杂问题）。困难问题虽然庞大，但可以通过理解、抽象和分层设计来解决，比如建造一座桥梁。而硬核问题，如设计一只能在各种地形中高速移动的猎豹，或识别人脸，其因果关系模糊，**无法通过简单堆叠零件或逻辑推演来“设计”求解。**人脸识别更多依赖**直觉**而非因果推理，这本质上是**神经计算**的产物。

解决硬核问题的正道不是设计，而是**发现。**这依赖于**模式识别**和**启发式。**解法不是精确计算出来的，而是通过**训练、迭代、试错和优化**“长”出来的。就像训练AI模型，我们从外部施加压力（如输入数据和反馈），让系统自己在反复变异、迭代和筛选中找到有效路径。这种“发现”的解法通常**实施速度快、灵活性极高，但其内部机制往往不可解释。**

复杂系统的一个核心特征是**涌现。**整体表现出的功能不能完全用构成它的部分来解释。大自然就是一个分层的涌现结构。每一层结构，如树叶的叶脉，都是解决特定问题（如传输水分、支撑叶片）的**物理抽象**，它们高效且适应性强。这些物理抽象不是被设计的，而是通过**元机制**（广义的自然选择/演化）在每一层反复筛选和迭代出来的。

面对复杂系统，我们需要结合**整体论**和还原论的视角。整体论关注系统的**宏观属性和抽象概念**，而非内部细节。例如，通过判断森林的温度和湿度来预防火灾，而无需知道具体的起火原因。这体现了**属性优于原因**的原则。理解复杂系统的共同**核心属性**（如自组织、非线性、反馈循环等） 能够帮助我们应对不确定和模糊的目标。

整体论思维的核心是**从外向内看、试错、重视反馈、以存活为底线。**这并非放弃理性，而是承认在复杂世界中，很多有效的方案是**“发现”而非“设计”**的。**学习还原论知识并非为了直接应用到复杂系统设计，更多是为了**类比**，从中提取模式和隐喻。AI作为整体论的重大成就，其功能是涌现的结果。

这种“发现”而非“设计”的理念，并非否定规划，而是将规划的重心从最终产物转移到“生成过程”本身。它要求我们拥抱不确定性，重视反馈，并以系统能否“存活”和持续迭代作为最终的检验标准 。在复杂系统中寻找高妙解法的正道，在于拥抱“发现”，而非执着于“设计”和寻找简单的因果关系。

为什么说人形机器人会率先实现大规模突破

一方面，专用型机器人的“小众”属性，注定很难撬动真正的产业飞轮。而相比之下，人形机器人应该会更有望率先实现大规模突破。这里的核心在于“规模效应”。目前市面上的专业机器人因为应用场景太窄，难以吸引持续的资金和技术投入，行业发展自然缓慢。而

人形机器人凭借对人类环境的高度适应性，能够胜任各种辅助角色，这为其打开了广阔的市场空间。只要一旦形成大规模生产，不仅能带来可观的利润，还会吸引更多投资和创新，形成正向循环，形成飞轮。 但另一方面，在新兴产业领域，尤其是在那些从零开始规划的“绿地”生产环境中，例如全新的超级工厂、深空探索基地或是极端作业场景，情况则大相径庭。在这些场景下，设计的首要目标是极致的效率、安全性和特定任务的完美执行，而非迁就人类的生理限制。因此，高度特化的非人形机器人，凭借其为特定任务定制的形态、传感器和执行器，将成为最优解。 这种双轨并进的格局，恰恰反映了技术发展在理想蓝图与现实约束下的必然选择。但是也决定了，更高的智能只能在人形机器人里面产生，人形机器人会率先实现大规模突破。

警惕大模型让人失去深度思考

最近看到一则国外课堂上的小插曲：一位老师情绪激动地感叹，如今的学生作业几乎都交给了大模型，鲜有人愿意静下心来独立思考。这一幕让我陷入沉思。每一次技术革命，似乎都在悄然间分割着人群——有人因新工具如虎添翼，收获颇丰；也有人被浪潮裹挟，失去了原本赖以安身立命的技能。

大模型的崛起让编程变得触手可及，写代码不再是少数人的专利。但随之而来的，是写代码这项能力本身的“贬值”。同样，大学里曾经精心设计、几代人不断打磨的课程体系和作业，原本意在锤炼学生的独立思考与问题解决能力，如今却可能沦为AI的“一键生成”，学生从“主动探索者”变成了“转手搬运工”，学习的本质——那种痛并快乐着的能力养成——被极大地稀释了。

对我来说，这其实也是一种警钟。大模型在我的日常学习和工作中，确实能帮我迅速梳理新闻、提炼论文要点、提出问题、甚至找到洞见和答案。但这真的让我成长了吗？我不禁自问。没有那种咬牙切齿、百转千回的思维挣扎，没有那种抽丝剥茧、灵光乍现的顿悟，智慧就像温水煮青蛙，悄无声息地滑过指缝。没有痛苦的学习，就像没有经历过亿万数据、万亿参数训练的大模型，怎会有真正的“聪明”？如果我们只是浅尝辄止地“懂了很多道理”，却未曾把这些知识转化为行动的指南，那我们的思维和认知，是否也会变得浮光掠影？

这让我想起学语言的过程。泛读泛听，固然能让你获得语感，快速了解一门语言的皮毛；但唯有精读精听、反复琢磨，才能真正把语言内化为自己的表达工具。同理，大模型可以带你迅速入门某个领域，帮你搭建知识的“脚手架”；但唯有系统化的深度思考、理论与实践的反复碰撞、归纳总结出属于自己的知识体系，才能真正让你成为领域里的行家里手。

有幻觉的GenAI也许才更有“人”味

在我们追问“如何消除大模型幻觉”的同时，是否忽略了幻觉本身正是AI最接近人性的地方？人类的记忆并不精确，往往只对最近的经历保有清晰的印象，遥远的往事则如迷雾般模糊。我们擅长用概率性和想象力补全自己的故事，哪怕那些细节未必真实。正因如此，人类的叙事才充满诗意、悬念与无限可能。

AI大模型的“幻觉”——即在信息不全时生成合理但未必真实的内容——其实与人类的认知机制惊人地相似。我们总希望AI像人一样思考、创作，但又渴望它绝对准确、永不出错。这种悖论，恰恰折射出我们对“智能”的复杂期待。

幻觉的本质，是概率的产物，是信息缺失时的最佳猜测。人类正是因为这种不确定性，才有了荡气回肠的爱情、经久不衰的故事、热血沸腾的英雄，才会创造聊斋、幻想科幻，才会对未知世界保持好奇。AI如果一味追求零幻觉，那它将失去想象力、创造力，沦为冷冰冰的检索工具。

当前生成式AI技术的两难：在医疗、法律等场景，幻觉是风险，但在创意、文学、科学假说生成等领域，幻觉却是灵感的火花。理想的AI，应能像人类一样，既能在需要时严谨克制，也能在合适的场合释放想象力。我们需要的不是“无幻觉AI”，而是“可控幻觉AI”——让AI的幻觉成为创新的引擎，而非误导的陷阱。

哲学上讲，幻觉是人类自我叙事的润滑剂，是我们理解世界、建构意义的方式。AI若能拥抱这种“有温度的不确定性”，或许才真正踏入“类人智能”的门槛。未来，AI与人类的共创，将建立在对幻觉的理解、驾驭与升华之上。

或许未来我们要问的不是“如何消除AI的幻觉”，而是“如何让AI产生有价值、有意义、可控的幻觉”，让AI不仅像人，更能和人类一起，探索未知、创造未来。

AI训练需要哪些“数据“?

伊利亚称互联网是AI的化石能源，但是现在(2025年初）可以用于训练模型的数据已经用完了。但是还有很多其他类型的AI训练没有足够的数据。比如机器人的训练。

Agent系统的演进 vs. 父母培养孩子的过程

Agent系统演进的过程就像父母培养孩子的理想状态：既有底线（workflow）和引导（协作Agent + 推理模型），又给足空间和资源（外部Tools+Resource）；既能保证安全和方向，又能激发创造和成长（自主Agent）。而孩子最终长大成人超越父母（Bitter Lesson）。

MCP Hub, MCP Store, MCP Registry, MCP Gateway, MCP Proxy

MCP Hub是MCP的官方仓库，用于存储MCP的协议文档和实现。

自动生成MCP Server

看起来大家都不约而同的希望通过大模型或Agent通过阅读MCP协议的文档和Server端应用能提供的服务，然后自动生成一个MCP Server。

A2A为什么这个时候出现

A2A (Agent-to-Agent)是Google公司提出的一个开源框架，我的理解A2A之于MAS，就像Kubernetes之于微服务系统，旨在通过多智能体之间的协作来提升MAS系统的能力和效率。A2A通过引入多个智能体之间的交互和协作，来实现更复杂、更高效的任务处理。之前的诸多MAS框架（如AutoGen、LangChain, CAMEL，MetaGPT等）都是使用自家的多Agent通信协议来实现的，他们之间是不能互通的，比如AutoGen的Agent找不到LangGraph的Agent，也不能与之通信，而A2A则强调了智能体之间的协作和信息共享。为打通各个MAS框架，A2A提供了一个统一的通信协议和交互方式，并还提供了一个统一的Agent Orchestrator和Agent 管理平台，允许用户在一个平台上管理和监控所有的智能体，想想K8s的集群管理，服务发现。并且A2A day0就支持MCP协议，这个就有点像K8S里面CxI（CNI，CSI, CRI等）的概念，A2A负责整个MAS的编排和管理，而MCP负责Agent和各种服务，Tool，信息源头之间的通信，就像K8s里面，Pod通过CNI访问网络，Pod通过CSI访问存储，Pod通过CRI使用不同底层容器技术。同理，在A2A里面，Agent通过MCP协议来和其他服务连接。

当然，如果你把你的Agent（或MAS）包装成MCP Server，那么它也可以与其他任何遵循MCP协议的Agent进行通信了。但是A2A的愿景是打通所有MAS框架，而MCP只是其中支持的一个协议罢了。A2A和MCP在某种程度上是互补的。2025年在各种Deep Research，（PC, Web）Operator，Claude Desktop，Manus出现后，俨然一副MAS大火的元年的架势，而google在这个时候推出支持MCP的MAS的编排框架A2A，显然是看到了这个趋势，并协同50家厂商一举占领市场用户的心智。在国内虽然也有ANP，都是从各方影响力来看差距很大，Google作为开源界的优等生兼超级大佬，市场的号召力和影响力是毋庸置疑的。

A2A：Google如何用"Kubernetes式思维"重新定义多智能体系统?

在2025年这个被业界称为"多智能体系统(MAS)元年"的时代，Google再次展现了其作为开源界超级大佬的前瞻性，推出了A2A(Agent-to-Agent)框架——这个可能彻底改变MAS生态的游戏规则改变者。

从碎片化到统一：A2A的颠覆性设计理念

想象一下Kubernetes对微服务世界的革命性影响，A2A对MAS领域带来的正是这种级别的范式转变。当前市场上的MAS框架——无论是AutoGen、LangChain、CAMEL还是MetaGPT——都像是一座座孤岛，各自使用专有的通信协议，导致不同框架的智能体根本无法相互发现和协作。这就像早期的容器编排系统，每家都有自己的解决方案，直到Kubernetes出现才统一了江湖。

A2A的核心创新在于它提供了一个**统一的通信协议和交互标准**，并配备了完整的**Agent Orchestrator**和**管理平台**。这相当于为MAS世界带来了K8s式的集群管理能力，让开发者能够在一个平台上管理和监控所有智能体，无论它们原本属于哪个框架。

MCP协议：A2A生态的"CNI/CSI/CRI"

特别值得关注的是A2A从Day 0就支持的MCP协议——这堪称MAS领域的"基础设施插件标准"。在Kubernetes中，我们有CNI(网络)、CSI(存储)、CRI(容器运行时)等标准接口；而在A2A生态中，MCP协议扮演着类似的角色，负责智能体与各种服务、工具和信息源之间的标准化通信。

这种设计的美妙之处在于它的**可扩展性**：任何将自己的Agent或MAS系统包装成MCP Server的实现，都能无缝接入A2A生态。但Google的野心显然不止于此——MCP只是A2A支持的众多协议之

一，其终极目标是成为连接所有MAS框架的"万能胶水"。

2025：MAS元年的天时地利

Google选择在2025年推出A2A绝非偶然。随着Deep Research、PC/Web Operator、Claude Desktop、Manus等创新产品的爆发式增长，MAS技术确实迎来了它的高光时刻。Google联合50家厂商共同推进A2A生态，这种"联盟式"打法不仅展现了其市场号召力，更是一种精心策划的生态占领策略。

相比之下，国内虽然也有ANP等类似尝试，但在影响力和生态建设上确实存在明显差距。作为开源界的"优等生"，Google再次证明了自己定义行业标准的能力——就像当年Android统一移动操作系统、Kubernetes统一容器编排一样，A2A很可能会成为MAS领域的事实标准。

未来展望：当每个Agent都成为A2A公民

A2A的出现预示着MAS发展将进入新阶段：

> **开发效率革命**：再也不用为不同框架的兼容性头疼
>
> **资源利用率提升**：跨系统的Agent协作成为可能
>
> **创新加速**：开发者可以专注于业务逻辑而非底层通信

这不禁让人想起Kubernetes早期的发展轨迹——从被质疑到被接受，再到成为行业标配。A2A是否也会沿着同样的路径发展？在MAS元年的背景下，答案很可能是肯定的。

作为技术人，我们或许正在见证一个新时代的开端——当A2A让每个智能体都能自由沟通协作时，真正的分布式人工智能才算是迈出了坚实的一步。Google这次又走在了前面，而我们要做的，就是准备好迎接这场由A2A带来的MAS生态大统一。

为什么MCP怎么火

MCP (Model-Context-Protocol)是xxx

火的原因有几个： 解决了几个痛点：当前的大模型不够聪明

OpenAI的Function calling不是行业规范，虽然它到目前为止是事实标准，但是其他的大模型对它的支持并不友好。而且这个由OepnAI完全掌控，会让别的大模型完全只能是follower，这显然不利于生态的健康发展。当新的接口出来后，其他大模型必须被动支持，时间上会滞后，甚至会被OpenAI的更新打乱节奏。

MCP是一个开源的标准，基于的技术JSON-RPC协议是非常成熟的通用规范，任何大模型理论上都已经支持它。MCP协议的变化性是在MCP Client和Server的实现上，而不是在协议本身，将来的演进更多是依赖于MCP协议而不是大模型。

MCP目前的短板是安全方面，MCP协议本身并不涉及安全性的考量。但是MCP Client和Server的实现是可以考虑安全的。比如，MCP Server可以只允许可信的Client连接，或者对某些敏感操作进行权限控制等。

将来的趋势： 短期趋势：MCP协议可以作为基于不同框架实现的Agent间，Agent和工具之间的通信协议，注意不是MAS框架。MCP协议的标准化和规范化将为AI Agent的开发和应用提供更多的可能性，特别是在多Agent协作、跨平台交互等方面。 长期趋势：MCP不再存在，因为当大模型都足够聪明的情况下，所有的协议都可以现学现用。而将来的协议可以是基于自然语言的协议，或者是大模型之间自己协商的。当然这个听起来比较科幻，仅仅用来开开脑洞。

> 目标是开源的规范：

RL强化训练的局限性

Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?

计算不可约性(Computational Irreducibility)

计算不可约性理论在AI Agent中的体现

当前AI Agent（如智能体、LLM驱动的工具型AI）常常需要借助外部工具（如搜索引擎、数据库、代码执行环境等）来完成复杂任务，尤其是当任务本身涉及大量不可预知、动态变化或信息量极大的情境时。

其理论解释是，在不可约性视角下，很多真实世界任务（如开放式问答、复杂推理、多步决策）本质上就是"不可约"的：没有一条简单的公式或神经网络能在内部一步到位地直接给出最终答案。 这也就意味着，AI Agent即使模型能力再强，也无法"内部化"所有世界知识和外部状态变化，只能通过调用外部工具/环境来"实际运行"所需的推理或数据检索过程，这与不可约性中"只能逐步模拟"高度一致。

以下是一些具体的例子：

搜索引擎调用：当Agent遇到新知识点时，无法仅靠训练参数推出答案，必须"查一查"——这就是对复杂系统不可约性的现实应对。

代码执行/环境交互：比如Copilot、GPT-4等Agent需要运行代码片段、与外部API交互，实际上是通过"外部模拟"来获得结果，绕不开计算不可约性带来的不可压缩性。

多Agent协作：多个Agent分工协作、彼此调用，也是在"分布式地"模拟一个不可约的复杂过程。

“计算不可约"对AI Agent系统设计的启示：

外部工具集成是必然趋势：既然不可约性普遍存在，AI系统设计时就应天然支持与外部世界的高效接口，而不是追求"全知全能的封闭模型”。

Agent的本质是"调度器"：智能体的核心价值，逐渐转向如何高效组织、调度外部资源和工具，提升整体推理效率，而不是仅靠内在模型参数"猜"出一切。

“计算不可约"在AI模拟城市的指导意义：

AI Agent或多智能体系统，正是通过模拟每个"市民”、“企业”、“交通工具"等微观个体的行为和交互，逐步推进城市状态演化。你无法通过简单的参数拟合或静态建模就预测出整个城市的未来状态，必须通过仿真（agent-based simulation, multi-agent system, reinforcement learning等）不断推进，才能发现潜在的涌现现象和复杂格局。这正是"计算不可约性"在工程实践中的最佳写照：现实问题太复杂，只能一步步算出来。

AI模拟城市的具体体现

城市交通仿真：交通流量、拥堵点、出行模式，只有通过交通微观模拟（如SUMO、MATSim等）才能真实再现，无法用公式直接预判。

城市政策实验：比如限号、调控、税收变化对经济和人口迁移的影响，只有通过多Agent模拟才能看到长期的动态反馈。

应急管理：灾害响应、疫情传播、资源调度等，都是高度不可约的过程，必须依赖AI或多Agent系统动态推演。

学习深度学习的理论和各种调优理论和工程方法有利于优化个人学习的方法论

深度学习是一群很牛的科学家试图用数学的方法来模拟大脑的神经网络，而大脑的神经网络是人类长期进化而来的，它具有很强的学习能力，能够通过少量的数据学习到很多知识，并且能够举一反三，融会贯通。 而对于普通个人来说，并不能熟练掌握高效的学习方法来极大挖掘人类大脑的能力。所以这群牛人从人脑学习方式挖掘出来一套方法论，通过实验成功应用到深度学习中，并取得了巨大的成功，而这套方法论其实可以反过来指导个人的掌握更优的学习方法，也就是说，深度学习的理论和方法论可以迁移到个人学习中，对优化个人学习的方法论有非常强的指导意义。这个看起来是个双向奔赴的过程，科学家从人脑的工作模式中得到启发，从而不断优化深度学习的方法论，而普通个人则可以借鉴深度学习的方法论，从而优化个人学习的方法论。

深度学习的训练，调优，微调，蒸馏，迁移学习，多任务学习，多模态学习，多Agent学习，等等，这些理论和方法论可以迁移到个人学习中，对优化个人学习的方法论有非常强的指导意义。 比如，训练讲究要优质数据（数据数量，质量，多样性，均衡性，等），这个和个人的学习很像，优质数据决定了模型学习的上限，而模型调优决定了模型学习下限，而掌握第一性原理所需要的优质知识，往往也是我们个人学习第一时间需要收集的。 训练中，数据集的划分，训练集，验证集，测试集，这个和个人的学习很像，个人的学习也需要分阶段，不同的阶段需要不同的数据集，比如，新手期，成长期，成熟期，探索期，等等。 训练中，超参数的调整，就像个人在学习中找到适合自己的学习方法，比如，学习率（学习速度），batch size（每次学习的内容量，比如，一次学习100个单词，还是1000个单词，每次半小时或1小时中间休息一下（比如番茄学习法）），epoch（每个周期，每个学科的学习次数），等等。 训练的优化，包括数据，模型，算法，这三个方面，而个人学习也是如此，个人学习需要找到适合自己的学习材料，学习方法，学习环境，学习伙伴，等等。 训练的优化，需要有目标，需要有评估标准，需要有优化方法，需要有反馈机制，需要有调整机制，需要有持续改进的机制，等等。 训练的优化，常常优化/满足指标框架与正交化原则可以协同工作。一旦满足某个指标达标，个人可以更"正交"地聚焦于使用特定的"旋钮"来提升优化指标，而不必过分担心对其他已达标指标的负面影响，比如，主要学科成绩，其他学科成绩，兴趣爱好，身体健康，心理健康，等等。这些指标之间是正交的，互不影响的。个人要做的是，找到主要学科的短板，优先改进，单个学科内部，找到短板，优先改进，比如语文，内部分为阅读，写作，古文，英语分为听说读写，等等。而各个知识点之间，又是正交的，互不影响的，个人要对所有知识点做到心中有

数，不遗漏，又能按部就班，循序渐进逐步掌握。不同的方法（调优），比如，快速阅读，深度学习（讲究举一反三，融会贯通，把好的题目做通做透），多模态学习（比如，图像，视频，音频，文本等），多Agent学习（比如，协作学习，竞争学习，等等），等等。

训练中，loss function的设计，这个和个人的学习中的反省，反思，总结经验教训很像，个人的学习也需要经常回顾之前的学习内容，效果，通过测试来评估学习效果，并根据测试结果调整学习方法。而要优化的目标也是多维度的，包括知识点的掌握，知识的灵活应用的程度，能否举一反三，能否创新，等等。 训练中，模型的选择，这个就像学习不同的学科，不同的学习方法之间既有大的相同之处，又有不同的细微之处，因为不同领域可能激活不同的大脑区域， 不同的学习方法（文字更需要抽象能力，图像（结构图，流程图，脑图等）更能极大压缩大量熟悉的信息。 而监督学习就像平时的考试，不断检验知识点是否都掌握，无监督学习，就像是自学，根据材料的上下文自组织学习和检验，迁移学习和强化学习更是举一反三，能在之前没有学习过的领域，快速掌握学习方法，实现跨界学习。

深度学习能学习任何东西吗?

数据 vs. 算法，哪个更重要?

长上下文 vs. 记忆

这里的思考主要关注大模型的长上下文的支持，以及大模型在记忆方面的能力。

大模型Scaling Law失效了吗?

当前Transformer-based大模型的局限性在哪里?

AI 多Agent系统的发展趋势

AI 多Agent系统的发展趋势主要体现在以下几个方面：

分享这篇文章

相关文章推荐

QwQ-32B Qwen推理..

本文介绍了深度求 ...

Deep Research ...

Deep Research 深度研究

Cursor AI 最佳实践：...

Cursor AI 最佳实践：提升编码...