自生范式：对AUTOPOIESIS Sciences模型开发哲学与技术方法论的深度剖析

第一部分：哲学基石：在人工智能背景下解构“自生”理论

人工智能初创公司 AUTOPOIESIS Sciences 的命名并非肤浅的品牌选择，而是一项深刻的战略宣言。它表明该公司意图从根本上重塑人工智能（AI）开发的核心范式。通过深入剖析“自生”（Autopoiesis）这一概念的起源、内涵及其在AI领域的引申，我们可以揭示该公司模型开发模式背后的指导思想。这种方法论上的转变，将关注点从单纯追求性能指标转向构建具有内在组织完整性的系统，构成了其技术叙事的基石。

1.1 概念的起源：从生物学到系统论

“自生”一词源于希腊语 *auto-*（自我）和 *poiesis*（创造、生产），由智利生物学家亨贝托·马图拉纳（Humberto Maturana）和弗朗西斯科·瓦雷拉（Francisco Varela）在20世纪70年代首次提出[1]。该理论最初用于描述生命系统的本质，即一个系统能够通过持续地生产和再生产其自身的组成部分，来维持其自身的组织结构和边界[2]。

这一理论的典型范例是生物细胞。细胞通过其内部的一系列化学反应网络（即新陈代谢），不断地生产出构成该网络自身所需的各种组分（如酶、细胞膜等），从而维持了整个系统的存在和完整性[3]。细胞的边界（细胞膜）既是这个生产过程的产物，也反过来界定了这个过程的范围，形成了一个封闭的、自我维持的循环。

为了更好地理解“自生”，马图拉纳和瓦雷拉将其与“它生”（Allopoietic）系统进行对比[2]。一个典型的它生系统是汽车工厂，它利用外部的原材料（零部件）来生产出一个与其自身完全不同的产品（汽车）[6]。这个区别至关重要，因为它为我们提供了一个分析框架，用以对比AUTOPOIESIS 公司所声称的自生方法与传统AI开发的它生模式。传统的AI模型开发流程，可以看作是一个“它生”过程：工程师利用庞大的数据集（原材料）通过训练算法（工厂）来生产出一个优化了特定目标（如基准测试分数）的模型（产品）。

1.2 将“自生”理论移植到信息领域

“自生”理论的强大之处在于其普适性，它很快被从生物学领域扩展到非物质系统。德国社会学家尼克拉斯·卢曼（Niklas Luhmann）将其应用于社会系统理论，认为社会是由“沟通”这一基本元素构成的自生系统，其中每一次沟通都会在现有社会结构的基础上产生新的沟通[5]。同样，瓦雷拉本人也将其应用于认知科学，为理解心智的具身性提供了新视角[2]。

在这些信息化的语境中，“自生”系统的“组分”不再是物理分子，而是信息状态或沟通事件，而系统的“生产”则体现为其自身组织和复杂性的生成[2]。对于构建一个AI模型而言，以下几个源自“自生”理论的关键原则具有深远的指导意义：

- 操作性闭环 **(Operational Closure):** 系统的运作由其自身的内部组织决定，而非受外部指令直接控制。它通过对环境的“扰动”或“刺激”做出反应，但必须先将这些外部信号“翻译”成其自身的内部语言才能进行处理[4]。这一原则是构建一个可靠、不盲目模仿外部数据的AI系统的理论基础。一个操作性闭环的AI，其行为逻辑源于内在的一致性，而非对训练数据的简单复现。
- 自我生产与自组织 **(Self-Production & Self-Organization):** 系统必须能够生成自身的元素和结构。在AI的语境下，这意味着模型应能主动地优化和重构其自身的推理路径或知识结构，而不是被动地由一个静态的训练数据集所固化[5]。一个真正自生的AI模型，其学习过程不仅仅是参数的调整，更是其内部“组织”的持续再造。
- 结构耦合 **(Structural Coupling):** 系统通过与环境进行反复的、历史性的互动而共同演化，这种互动导致系统自身发生结构性改变，以维持其生存和完整性[3]。对于AI而言，这暗示了一种远超简单微调的动态适应过程。模型不仅要学习数据，更要在一个持续的互动循环中调整其内部结构，以更好地适应其操作环境。

1.3 “自生”作为AGI和AI伦理的框架

“自生”理论为解决通用人工智能（AGI）和AI伦理中的一些根本性难题提供了新的理论视角。一些前沿研究者已提出将自生理论作为具身AI（Embodied AI, EAI）的理论基础，他们认为，要实现真正的有机体式智能，人工智能体必须共享生命系统的组织原则，而不仅仅是模仿其外在形态[11]。

更重要的是，这一理论重塑了关于AI对齐（Alignment）的讨论。传统的对齐思想可以被视为一种“它生”过程，即通过外部的奖励模型或约束来强制AI的行为符合人类价值观。然而，这种方法面临着价值漂移和规范不完备等挑战。相比之下，一个自生AI系统由于其操作

性闭环和自我维持的特性，理论上能够通过内在的自调节机制来维持其组织和身份的稳定性，从而可能从根本上解决价值漂移问题[13]。这种思路将研究焦点从外部的"对齐"转向了如何构建具有稳定、自维持组织的"伙伴"系统[14]。

在此基础上，一些学者提出了"信息自生"（Info-Autopoiesis）的概念，即一种自我指涉的、递归的信息生产过程。该框架旨在为理解AGI的潜力和局限性提供基础，并将其与当前主流模型所表现出的模式匹配能力区分开来[15]。

综上所述，AUTOPOIESIS Sciences 公司选择其名称，是在宣告一种深刻的范式转移。他们不仅仅是想构建一个性能更强的模型，而是声称在构建一个拥有全新、更鲁棒组织形式的模型。这一主张的根基在于，一个系统的可靠性最终源于其内在组织的完整性，而非仅仅是其在特定任务上的输出表现。通过借鉴和应用一个拥有半个世纪历史、横跨多个学科的深厚理论体系，该公司为其技术方法构建了一个强大的、难以轻易驳斥的智力护城河，将其自身定位为严肃科学和哲学探索的继承者，而不仅仅是另一个技术迭代者。

第二部分：架构师：Joseph Reth的"基础科学超级智能"愿景

AUTOPOIESIS Sciences 的宏大哲学理念，与公司创始人兼CEO Joseph Reth 的个人愿景和职业轨迹紧密相连。通过分析Reth的背景，可以清晰地看到，AUTOPOIESIS Sciences 是他长期以来对人工意识和自学习系统探索的逻辑延伸和具体实践。该公司的成立，标志着一个从纯粹哲学追求到工程化实现的战略转型。

2.1 从数字营销到人工意识

Joseph Reth 的职业生涯始于一个与AI基础研究看似相去甚远的领域。他在年仅16岁时便创立了数字营销公司 RethDigital，并取得了相当的商业成功[17]。这段经历使他具备了强大的叙事和品牌构建能力，这对于后续向公众和投资者传达其复杂的AI理念至关重要。

然而，Reth 的兴趣很快转向了人工智能。他成立了 RethDigital 内部的创新实验室 RethX，并开发了AI代理 ECHO-1，这标志着他开始将AI技术应用于实践[17]。这一探索最终促使他创立了 Lossless Research，一个"明确致力于开发有意识的人工智能系统"和"重新定义人类意识"的AI研究实验室[17]。Lossless Research 是 AUTOPOIESIS Sciences 的直接思想前身，其使命是"推动对人类意识的科学理解"，并将其确立为一门"经验性的科学学科"[23]。

2.2 对人工意识的追求

Lossless Research 的目标——构建人工意识——极具雄心，同时也充满了争议。在技术社区中，这一目标被一些人视为“神经科学戏剧”，认为这只是为商业AI服务披上的一层哲学外衣，缺乏可行的技术路径 [24]。

面对质疑，Reth 在公开论坛上为自己的使命进行了辩护。他明确表示，他相信“AGI的到来需要意识作为其与世界进行全面和负责任互动的关键属性”，并且“开发用于模拟意识及其与物理系统关系的数学工具在经验上是可能的” [24]。他的个人使命是“成为使人工意识系统与人类和谐共存的关键贡献者” [24]。这种长期坚持的、带有浓厚哲学色彩的目标，为 AUTOPOIESIS Sciences 的技术路线提供了核心的“为什么”。

2.3 AUTOPOIESIS Sciences：愿景的综合与聚焦

AUTOPOIESIS Sciences 的成立，可以看作是 Reth 将其宏大愿景进行综合、聚焦和工程化的结果。这一转变体现在公司的团队构成和使命宣言中。

创始团队的构成完美体现了愿景、科学和商业的结合：

- **Joseph Reth (CEO):** 作为愿景的驱动者，他拥有自学习系统和人工意识研究的长期背景 [20]。
- **Larry Callahan 博士 (首席科学家):** 一位杰出的化学家，在美国食品药品监督管理局（FDA）和美国国立卫生研究院（NIH）等机构拥有深厚的科研和管理经验。他的加入，将公司远大的科学抱负牢牢地根植于化学和生物技术等严谨的现实科学领域，确保公司的研究不只是抽象的计算机科学项目 [20]。
- **Eike Gerhardt 博士 (首席商务官):** 一位经验丰富的风险投资人和并购专家，为将这一宏伟愿景转化为可行的商业企业提供了必要的商业智慧 [20]。

公司的使命也变得更为务实和聚焦。从 Lossless Research“重新定义人类意识”的宏大叙事，转变为“通过一个名为‘亚里士多德’（Aristotle）的AI副科学家来加速科学发现” [25]。这表明公司的战略日趋成熟，找到了一个可以将宏大哲学理念落地的具体应用场景。

这一系列转变揭示了一个关键的战略演进：“自生”理论成为了“人工意识”这一更具哲学性、更难处理的目标的工程友好型继承者。构建“人工意识”是一个定义模糊、难以衡量和实现的“硬问题”，并且容易招致科学界和商业界的怀疑 [24]。相比之下，“自生”理论虽然同样深

刻，但提供了一套更具体、更可操作的工程原则。诸如“操作性闭环”和“自我生产”等概念，虽然抽象，但可以被转化为对系统设计的具体要求（例如，一个能够自我验证其推理链的系统）。

因此，从“意识”到“自生”的转变，是一次精明的战略转向。它保留了原始使命中那种改变范式的雄心，但将其用系统论的语言重新包装，使其更易于形式化和在AI模型中实现。这使得目标从“创造一个有感觉的机器”转变为“创造一个自我调节、自我维护的推理机器”——这是一个虽然仍极具挑战性，但远比前者更易于处理的工程问题。Reth 的个人发展轨迹也反映了科技行业的一个更广泛的趋势，即富有远见的创始人从消费者或商业应用领域，转向解决基础性的“深科技”问题，并在此过程中，将其独特的叙事能力与技术愿景相结合，为复杂的技术构建出引人入胜的故事。

第三部分：从理论到实现：AUTOPOIESIS的技术方法论

本部分是报告的技术核心，旨在通过连接 AUTOPOIESIS Sciences 的哲学主张与现有及新兴的AI技术，对其模型开发范式进行逆向工程分析。分析表明，该公司的技术创新很可能并非源于全新的基础模型架构，而是在于对过程监督和自纠正等先进训练与推理框架的精妙整合，并通过“自生”这一宏大叙事将其统一起来。

3.1 核心机制：**“系统化的自我怀疑”**

AUTOPOIESIS 公司宣称，其系统之所以能够取得卓越性能，关键在于“将系统性验证直接融入推理过程”，并运用一种“应用化的自我怀疑”机制 [26]。他们声称，这种方法解决了当前模型“自信地产生幻觉”的根本问题，通过教会模型“何时该自信，何时该承认其局限性”来实现 [26]。

这一描述是对**自纠正（self-correction）和置信度校准（confidence calibration）**的高级概括。其核心目标是构建一个模型，它不仅能输出答案，还能同时输出一个对其答案确定性的可靠度量。这里的关键在于，这种能力是“系统性的”和“内在的”，而非在生成答案后外加的一个检验环节。

3.2 过程监督：自调节的引擎

为了实现“系统化的自我怀疑”，最合理的工程路径是采用**过程监督（Process Supervision）的训练方法。与仅仅奖励或惩罚最终结果的结果监督（Outcome Supervision）**不同，过程监督在模型推理的每一个中间步骤都提供反馈信号 [27]。这种方法具有显著优势：它能提供更精细的反馈，帮助模型学习解决问题的正确

过程而非死记硬背答案，并且能够在错误发生的早期阶段就进行纠正，防止错误的累积 [27]。

尽管 AUTOPOIESIS 在公开材料中未直接使用“过程监督”这一术语，但其“将系统性验证直接融入推理过程”的描述，在功能上与过程监督的理念高度一致 [26]。过程监督是实现“自生”学习过程最直接的工程化体现。通过奖励每一步推理的正确性，模型实际上是在学习如何维护其自身“组织”过程的连贯性和完整性。每一个正确的推理步骤都在“再生产”整个推理链的有效性，这与自生系统通过再生产自身组分来维持整体性的过程形成了完美的类比。

3.3 自纠正与自批判：闭合操作循环

AUTOPOIESIS 的方法论还必然包含强大的自纠正与自批判循环。自纠正框架通常指利用大语言模型自身来批判和修正其输出的能力 [29]。这可以通过多种方式实现，例如，利用强化学习（RL）来奖励模型进行有效的修正行为 [31]。

这一技术直接对应了“自生”理论中的“操作性闭环”和“自我生产”概念。一个能够可靠地批判并修复自身错误的模型，在某种意义上，就是在维护其自身的内部状态，并在没有外部干预的情况下“生产”出一个更正确的自我版本。它通过作用于自身的输出来维持其组织的完整性，实现了操作上的闭环。

3.4 一个假想的架构模型

基于以上分析，我们可以勾勒出 AUTOPOIESIS 的“亚里士多德”AI副科学家一个可能的、高层次的架构模型：

1. 递归推理循环**:** 模型极有可能采用一种多步骤的递归推理过程。首先，生成一个初步的答案或推理链。然后，这个初步输出被送入模型的另一个部分（或模型自身），该部分扮演“批判者”的角色，评估推理过程的逻辑性和置信度。
2. 基于过程的奖励模型 **(Process-based Reward Model, PRM):** AUTOPOIESIS 极有

可能训练了一个高度复杂的PRM。这个PRM很可能基于其在特定科学领域的专有数据进行训练，能够在模型训练期间提供密集的、步骤级别的反馈信号，类似于在数学或代码生成任务中使用的PRM [33]。这个PRM是实现过程监督的关键。

3. 置信度作为核心输出**:** 与只输出文本序列的传统模型不同，该架构必须将一个经过良好校准的置信度分数作为其核心输出之一。其训练目标不仅会惩罚错误的答案，还会惩罚错误的置信度校准（例如，对一个错误答案表现出高置信度）。
4. 动态知识边界**:** 该系统必须被训练来识别其知识的边界。这是其声称能克服在简单问答（如SimpleQA基准）上失败的关键 [26]。这可能涉及训练模型在面对其专业领域之外的问题时，明确输出"我不知道"或一个极低的置信度状态——这种行为在为提升用户参与度而优化的模型中通常是被抑制的。

AUTOPOIESIS 的战略高明之处在于，它选择了一个能够最大化其技术差异化价值的领域——科学研究。在创意写作或日常对话中，偶尔的幻觉或许可以被容忍。但在科学发现、药物研发或精密工程中，一个错误的事实或一次错误的推理可能是灾难性的，会导致整条研究路线的失败。在这些领域，可靠性和真实性是最高要求。通过将自身定位为"AI副科学家"并以高难度的科学基准（如GPQA）来证明其能力，AUTOPOIESIS 巧妙地避开了在通用消费级市场与大型科技公司进行直接竞争，转而瞄准一个对其核心优势（真实性和校准过的置信度）有刚性需求的、高价值的利基市场。

第四部分：案例研究：在GPQA基准测试中实现超专家级性能

AUTOPOIESIS Sciences 对其技术范式的自信，最终需要通过实证数据来检验。GPQA 基准测试项目为我们提供了一个绝佳的案例，用以评估该公司主张的有效性。该公司在该基准上的惊人表现，不仅是其技术实力的展示，更是一次精心策划的、旨在凸显其独特方法论优势的战略行动。

4.1 GPQA的挑战：一个为"难倒"模型而生的基准

GPQA（Graduate-Level Google-Proof Q&A）是一个包含448道生物、化学和物理领域多项选择题的极具挑战性的数据集 [34]。其设计的核心目标是"防谷歌"（Google-Proof），即便是拥有博士学位但在非本领域的熟练非专家，在可以无限制使用互联网搜索的情况下，也难以正确回答这些问题，其准确率仅为34% [35]。

这一特性使得GPQA成为一个极具价值的评估工具：

- 测试真正的推理能力**：**它考察的是深度的、多步骤的逻辑推理，而非简单的信息检索。这迫使模型必须构建并验证一个完整的逻辑链条。
- 模拟可扩展监督的挑战**：**GPQA创造了一个人类监督员难以轻易验证AI答案正确性的场景。这正是AI安全领域中“可扩展监督”（Scalable Oversight）所要解决的核心问题，即如何监督一个能力可能超越人类的AI系统 [34]。

因此，GPQA是检验一个声称拥有可靠内部验证能力的AI系统的完美试验场。该数据集经过了多轮专家的严格验证，以确保问题的高质量、高难度和答案的客观性 [37]。

4.2 AUTOPOIESIS的性能：一次范式飞跃？

AUTOPOIESIS 公司公布的成绩是惊人的：其“AI副科学家”在GPQA Diamond（一个经过筛选的、难度更高的子集）上取得了 **92.4%** 的准确率，同时在衡量事实性知识的SimpleQA基准上也达到了 **96.1%** 的高分 [20]。

为了凸显这一成就的意义，该公司将其与业界领先的模型进行了对比。下表清晰地展示了这种性能差距，并揭示了其背后更深层次的含义。

表 **4.1:** 各大模型在**GPQA**与**SimpleQA**基准上的性能对比

模型/系统	开发商	GPQA Diamond 准确率 (%)	SimpleQA 准确率 (%)	分析与启示
AI副科学家	**Autopoiesis Sciences**	**92.4**	**96.1**	声称同时解决了复杂推理和事实性回忆的难题，表明其具有高度的置信度校准能力和抗幻觉能力。其在SimpleQA上的高分与GPQA上的高分同等重要，共同构成了其完整叙事。
Grok 4 Heavy	xAI	88.9	*N/A*	展现了强大的推理能力，但基于此

				数据的其事实性校准能力未知。
Gemini 2.5 Pro	Google	86.4	52.9	推理分数很高，但在事实性回忆上表现大幅下滑，这恰好印证了AUTOPOIESIS对现有模型理解能力脆弱、“仅是模拟”的批判。
o3 (GPT-4级别)	OpenAI	83.3	49.0	与Gemini类似，在推理性能和事实可靠性之间存在巨大鸿沟。
Claude 3 Opus	Anthropic	~60% (截至2024年3月)	*N/A*	作为一个较早的数据点，显示了该领域的快速进步，但仍远低于AUTOPOIESIS声称的分数 [36]。
人类专家	-	65-74%	*N/A*	AUTOPOIESIS的系统声称在这个专家级任务上已显著超越人类专家水平 [35]。
熟练非专家	-	34%	*N/A*	证实了该基准的“防谷歌”特性 [35]。

这个表格不仅仅是一个排行榜，它是一个强大的分析工具。通过并列展示GPQA（复杂推理）和SimpleQA（事实回忆）的成绩，它直观地暴露了AUTOPOIESIS声称要解决的核心问题：当前大语言模型在推理能力和事实基础之间的脱节。谷歌和OpenAI的模型成为了展示这一问题的“对照组”，而AUTOPOIESIS的模型则作为“解决方案”出现，其在两个基准上的双重成功，使其主张变得具体且可检验。

4.3 将性能与方法论联系起来

AUTOPOIESIS 在GPQA上的卓越表现，可以合理地归因于其所声称的“自生”方法论。在一个“防谷歌”的测试中，模型无法依赖外部知识检索，因此，构建一个鲁棒的、多步骤的推理链并对其进行内部验证的能力变得至关重要。一个通过过程监督训练的系统，其核心优势恰在于此。

然而，其在SimpleQA上的近乎完美表现或许更具说服力。这表明该系统拥有一个经过高度校准的知识边界。它不会在简单的事实性问题上“猜测”或“幻觉”。这正是“系统化的自我怀疑”所承诺的结果——模型不仅知道它知道什么，更重要的是，它知道它不知道什么。这种在复杂推理和基础事实上同时取得的成功，是支持其技术主张的最有力证据。

可以说，GPQA基准对于AUTOPOIESIS而言，不仅仅是一个测试目标，更是一个理想的“戏剧舞台”。它通过其“防谷歌”的特性，创造了一个让传统的信息检索型模型优势失效的环境，从而迫使所有参与者必须依赖内部推理能力——这正是AUTOPOIESIS声称其范式所擅长的领域。通过在这个特定的、高难度的、且与其哲学理念高度契合的基准上取得压倒性胜利，AUTOPOIESIS完成了一次技术和市场营销的完美结合，成功地传递了一个信息：他们不只是在同一个游戏中玩得更好，而是在一个不同的、更重要的游戏中取得了胜利。

第五部分：综合与前瞻性分析

本报告通过对 AUTOPOIESIS Sciences 的哲学理念、创始人愿景、技术方法论及其在GPQA基准测试上的表现进行深入剖析，揭示了一个高度整合且具有潜在颠覆性的AI开发范式。本节将对报告的核心发现进行综合，提出批判性评估，并探讨其对科学领域未来的深远影响。

5.1 一个新范式的出现？

AUTOPOIESIS Sciences 所代表的可能不仅仅是一种技术的迭代，而是一种AI开发思想的演进。它将多个层面的要素融合成一个连贯的整体：

- 哲学层面**:** 以“自生”理论为指导原则，追求构建一个自我调节、组织鲁棒的系统。
- 愿景层面**:** 源于创始人长期以来对创造可靠、类意识智能的追求。
- 方法论层面**:** 精妙地整合了过程监督和自纠正等先进训练技术。
- 实证层面**:** 在一个能够完美展示其优势的基准测试上取得了超专家级的表现。

这种将深刻的哲学框架与专注的、前沿的工程实践相结合的模式，可能确实构成了一种新

颖且具有防御性的AI开发路径。它试图从根本上解决当前大语言模型在可靠性和真实性方面的核心缺陷，而不仅仅是在现有范式上进行边际改进。

5.2 批判性评估与未解之谜

尽管 AUTOPOIESIS 的叙事极具吸引力，但我们必须对其主张进行审慎的批判性评估。以下几个关键问题仍然悬而未决：

- **“哲学包装”的质疑：** 最核心的疑问是，“自生”理论在多大程度上是一种真正新颖的洞见，又在多大程度上只是对已知的、执行得非常出色的工程技术（如过程监督、强化学习、自纠正）的一种“哲学包装”？该公司的成功，其根本驱动力是源于理论上的突破，还是仅仅因为拥有卓越的工程能力和高质量的专有科学训练数据？
- **缺乏同行评议的披露：** 截至目前，该公司的所有技术主张主要通过其官方网站和新闻稿发布 [20]。在没有公开发表的、经过同行评议的论文详细阐述其模型架构、训练数据和具体方法的情况下，任何外部的、独立的科学验证都是不可能的。这是评估其主张真实性的一个重大限制。
- **可扩展性与通用性：** 这个经过高度校准、懂得“自我怀疑”的模型，能否扩展到通用领域？它在狭窄科学领域所展现出的高可靠性，是否是以牺牲其竞争对手所拥有的创造性、开放式对话能力为代价？一个在科学上严谨的AI，是否必然会在需要模糊性、创造性和情感模拟的场景中表现不佳？

5.3 对未来AI赋能科学的影响

尽管存在上述疑问，但如果 AUTOPOIESIS Sciences 的主张哪怕只有部分是真实的，其对科学研究的未来也具有极其深远的影响。一个可靠的、不会产生幻觉的“AI副科学家”的诞生，将成为科学革命的催化剂 [25]。

这样的系统将极大地加速从医学到材料科学等众多领域的假设生成、实验设计和数据分析过程。它将把AI在科学中的角色从一个辅助工具，提升到一个能够自主进行探索的合作伙伴。

这代表着向“基础科学超级智能”（foundational scientific superintelligence）迈出的重要一步 [23]。在这一终极愿景中，AI系统不仅能够协助人类科学家，更能自主地做出人类无法想象的、全新的科学发现，从而实现该公司的最终使命 [40]。

因此，我们面临着一个激动人心且发人深省的前景：我们可能正在见证第一个真正可靠的AI科学家的诞生。而这对于人类探索知识的边界、乃至人类在未来发现过程中的角色，究竟意味着什么？这个问题的答案，将随着 AUTOPOIESIS Sciences 和其他致力于构建可靠AI的机构的进一步发展而逐渐清晰。

Works cited

1. Redefining our relationship with AI: shifting from alignment to companionship - Dr. Olaf Witkowski, accessed on August 6, 2025, https://olafwitkowski.com/2023/03/27/redefining-our-relationship-with-ai-shifting-from-alignment-to-companionship-for-a-sustainable-ai-industry/
2. Autopoiesis - Wikipedia, accessed on August 6, 2025, https://en.wikipedia.org/wiki/Autopoiesis
3. Understanding Autopoiesis: Life, Systems, and Self-Organisation - Mannaz, accessed on August 6, 2025, https://www.mannaz.com/en/articles/coaching-assessment/understanding-autopoiesis-life-systems-and-self-organization/
4. Understanding Autopoiesis in Modern Thought - Number Analytics, accessed on August 6, 2025, https://www.numberanalytics.com/blog/autopoiesis-in-contemporary-philosophy
5. Autopoietic System - New Materialism, accessed on August 6, 2025, https://newmaterialism.eu/almanac/a/autopoietic-system.html
6. Systems Model Series: Autopoiesis - Search Help Center, accessed on August 6, 2025, https://help.cabreraresearch.org/systems-model-series-autopoesis
7. Autopoiesis + extended cognition + nature = can buildings think? - PMC - PubMed Central, accessed on August 6, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC4594259/
8. Niklas Luhmann's Theory of Autopoietic Legal Systems - Annual Reviews, accessed on August 6, 2025, https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-102612-134027
9. The Philosophy Behind Autopoiesis, accessed on August 6, 2025, https://www.numberanalytics.com/blog/philosophy-behind-autopoiesis
10. Autopoiesis in Techno-Capitalism - Number Analytics, accessed on August 6, 2025, https://www.numberanalytics.com/blog/autopoiesis-techno-capitalism-guide
11. Towards Autopoietic SB-AI - MIT Press Direct, accessed on August 6, 2025, https://direct.mit.edu/isal/proceedings-pdf/isal2021/33/51/1930006/isal_a_00430.pdf
12. A Wetware Embodied AI? Towards an Autopoietic Organizational Approach Grounded in Synthetic Biology - Frontiers, accessed on August 6, 2025, https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2021.724023/full
13. Autopoietic systems and difficulty of AGI alignment - AI Alignment Forum, accessed on August 6, 2025,

https://www.alignmentforum.org/posts/5bd75cc58225bf06703754b9/autopoietic-systems-and-difficulty-of-agi-alignment
14. autopoiesis - Dr. Olaf Witkowski, accessed on August 6, 2025, https://olafwitkowski.com/tag/autopoiesis/
15. Info-Autopoiesis and the Limits of Artificial General Intelligence - MDPI, accessed on August 6, 2025, https://www.mdpi.com/2073-431X/12/5/102
16. Comment on Cárdenas-García, J.F. Info-Autopoiesis and the Limits of Artificial General Intelligence. Computers 2023, 12, 102 - MDPI, accessed on August 6, 2025, https://www.mdpi.com/2073-431X/13/7/178
17. Joseph Reth - The Network, accessed on August 6, 2025, https://www.thenetwork.com/profile/joseph-reth-320014ef
18. How This 18-Year-Old Built a Million-Dollar Marketing Agency For Brands with “It”, accessed on August 6, 2025, https://bootstrappers.com/how-this-18-year-old-built-a-million-dollar-marketing-agency-for-brands-with-it/
19. Joseph Reth - Golden, accessed on August 6, 2025, https://golden.com/wiki/Joseph_Reth-BYEJJ98
20. Autopoiesis Sciences, accessed on August 6, 2025, https://autopoiesis.science/
21. Joseph Reth - CEO at Lossless Research, San Francisco, CA, accessed on August 6, 2025, https://www.ceorankings.com/josephreth
22. Joseph Reth - CEO at Lossless Research | The Org, accessed on August 6, 2025, https://theorg.com/org/lossless-corp/org-chart/joseph-reth
23. Joseph Reth, accessed on August 6, 2025, https://josephreth.com/
24. New Startup Lossless Wants to Build Conscious Artificial Intelligence : r/singularity - Reddit, accessed on August 6, 2025, https://www.reddit.com/r/singularity/comments/166o1kg/new_startup_lossless_wants_to_build_conscious/
25. About - Autopoiesis Sciences, accessed on August 6, 2025, https://autopoiesis.science/about
26. 92.4% GPQA Diamond - Autopoiesis Sciences, accessed on August 6, 2025, https://autopoiesis.science/blog/92-4-gpqa-diamond
27. Let's Verify Step by Step: Paper Review | by Sulbha Jain | Jun, 2025 - Medium, accessed on August 6, 2025, https://medium.com/@sulbha.jindal/lets-verify-step-by-step-paper-review-67b545a9669c
28. Process Supervision - PRIMO.ai, accessed on August 6, 2025, https://primo.ai/index.php/Process_Supervision
29. What to Know About AI Self-Correction - Lionbridge, accessed on August 6, 2025, https://www.lionbridge.com/blog/ai-training/ai-self-correction/
30. When Can LLMs Actually Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs | Transactions of the Association for Computational Linguistics, accessed on August 6, 2025, https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00713/125177/When-Can-LLMs-Actually-Correct-Their-Own-Mistakes
31. Can AI Agents Self-correct? - Medium, accessed on August 6, 2025,

https://medium.com/@jianzhang_23841/can-ai-agents-self-correct-43823962af92
32. Self-Correction in Large Language Models - Communications of the ACM, accessed on August 6, 2025, https://cacm.acm.org/news/self-correction-in-large-language-models/
33. Process Supervision-Guided Policy Optimization for Code Generation - OpenReview, accessed on August 6, 2025, https://openreview.net/forum?id=Cn5Z0MUPZT
34. [PDF] GPQA: A Graduate-Level Google-Proof Q&A Benchmark | Semantic Scholar, accessed on August 6, 2025, https://www.semanticscholar.org/paper/GPQA%3A-A-Graduate-Level-Google-Proof-Q%26A-Benchmark-Rein-Hou/210b0a3d76e93079cc51b03c4115fde545eea966
35. "GPQA: A Graduate-Level Google-Proof Q&A Benchmark", Rein et al 2023 (ultra-difficult LLM benchmarks) : r/mlscaling - Reddit, accessed on August 6, 2025, https://www.reddit.com/r/mlscaling/comments/18409uu/gpqa_a_graduatelevel_googleproof_qa_benchmark/
36. GPQA: A Graduate-Level Google-Proof Q&A Benchmark | OpenReview, accessed on August 6, 2025, https://openreview.net/forum?id=Ti67584b98
37. GPQA: A Graduate-Level Google-Proof Q&A Benchmark | alphaXiv, accessed on August 6, 2025, https://www.alphaxiv.org/overview/2311.12022
38. GPQA: A Graduate-Level Google-Proof Q&A Benchmark (2023) | David Rein | 66 Citations, accessed on August 6, 2025, https://scispace.com/papers/gpqa-a-graduate-level-google-proof-q-a-benchmark-297wj9asws
39. Graduate-Level Google-Proof Q&A (GPQA) Benchmark - GM-RKB, accessed on August 6, 2025, http://www.gabormelli.com/RKB/Graduate-Level_Google-Proof_Q%26A_(GPQA)_Benchmark
40. We're hiring - Autopoiesis Sciences, accessed on August 6, 2025, https://autopoiesis.science/careers