Ensoul：一种通过进化能静态网络创建自组织智能超低功耗系统（SOULS）的框架。

Ty Roachford 佛罗里达大西洋大学机器感知与认知
机器人学（MPCR）实验室
Ty@TyfoodsForThought.com

arXiv:2304.13863v1 [cs.AI] 26 Apr 2023

摘要

Ensoul 是一种框架，旨在通过结合使用能量稳态（能静态）回路网络及其嵌套结构，以及开放式进化技术，来创建能够创造更多技术的技术。采用这种方法开发的生成式技术，既可作为由热力学驱动的复杂系统的简单而富有启发性的模型，也可作为新型技术的强大来源。“自组织智能超低功耗系统”（SOULS）这一术语既恰当地描述了这种生成式技术所产生的技术，也描述了生成式技术本身。该术语旨在体现这类技术的抽象性质，即它们独立于其所嵌入的基底。换言之，SOULS 可以是生物学的、人工的，也可以是混合形式的。

动机与概要

本文旨在串联一些主题，这些主题有望帮助我们深入理解有关生命与智能的基本问题，同时也有助于阐明本文采用特定方法来创造生成式技术的动机。本文采用较为宏观的视角，因此省略了某些主题的具体技术细节。这样做有几个原因。首先，本文力求简洁。理解各主题之间的联系，并不需要掌握每个主题的许多具体细节。其次，这样也便于没有相关专业背景的读者更轻松地跟上论述脉络，从而尽可能提高本文的可读性。希望深入了解的读者可以参阅文中列出的各类参考资料。

推动我们采用演化能静态网络方法来构建生成式技术的主要问题如下：

- 自然演化系统中的各个组成部分如何汇聚起来，为共同利益而协同工作？换句话说，自组织是如何运作的？
- “嵌入式智能体”——也就是说，由其所处环境中的同一种“物质”构成的智能体——如何在逻辑不确定性下做出决策（即只能在有限的时空范围内感知环境，且处理能力有限）？
- 如何在计算机模拟中捕捉自然界的复杂性，而不必随机搜索可能模型的空间，也不必仅依靠数学方法来创建、调整和运行参数繁多且在生物学上准确的模型？

因此，本文综合讨论并拟进一步探索的主要主题包括：

- 作为强大模型的能量稳态（能稳态）回路
- 自由能原理（FEP）作为对此类模型的有效描述

- 通过能量稳态循环构成的能量封闭系统实现开放式演化
- 组装理论可作为对这类封闭系统的有效描述。

具体而言，本文分为以下八个部分：

1. **“目的论、TAME 与 SOULS”** 部分阐述了为何将系统视为具有“目标”是有益的，这一观点受到自由能原理、控制论以及“普遍心智的技术路径”（TAME）框架的启发。在此背景下，文中还解释了“SOULS”一词，指出它是对这类不依赖特定基底的现象的理想描述。

2. **“开放式演化与组装理论”** 部分介绍了“开放式进化”的实际含义，以及它与进化编程领域目标之一的关系。该部分还讨论了开放式进化与组装理论之间的联系，并从更宏观的视角审视组装系统。

3. **”IOS 幻象：以能静态网络为建模提供动机”** 部分阐述了为何能量稳态循环、网络和嵌套结构是建模 SOULS 的理想抽象，相较之下，“输入输出”系统方法并非如此。

4. **“能静态网络”**部分通过简单的代码说明，讨论能静态循环的工作方式，并概述如何将其专门用于构建神经网络模型；这种模型的泛化最终促成了抽象能静态网络的诞生。

5. **“能静态网络的进化”**部分是本文的主要贡献。该部分描述了如何通过使用包含能静态循环的能量封闭系统来实现开放式进化；这些能静态循环会自然地相互作用，形成能静态网络和巢。如同“能静态网络”部分一样，本文在不规定具体实现细节的前提下，尽可能提供了详细信息。

6. **“热力学计算与能静态网络”**部分解释了热力学计算的含义，并提出能静态循环、网络和巢或许可以作为一种可扩展的通用模型系统，用于推动热力学计算的发展。

7. **“驯服 SOULS”**部分将能静态网络解释为实施合作与竞争动力学的系统；这些动力学所产生的后果驱动着行为。该部分还讨论了理解和控制此类动力学的方法。具体而言，文中讨论了“受控变量测试”，并概述了“变量可控性测试”和“控制开关测试”的总体框架。

8. **“讨论：研究方向、推测与哲学”** 提出了可能的研究方向，推测了这类研究的潜在益处，最后讨论了深入科学知识所带来的哲学意涵。

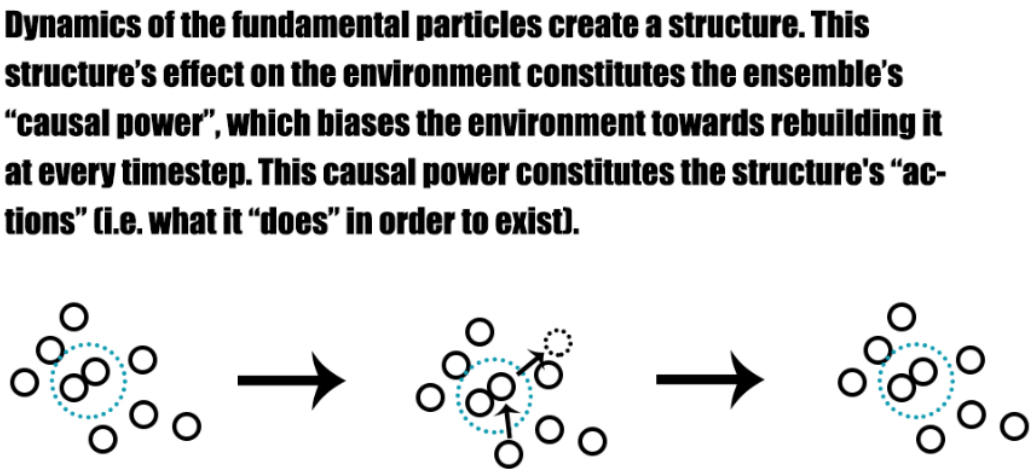
所有章节都采用了更具对话感、更像“讲故事”的形式，而非“纯粹”的学术论文风格。关于实际实验的经验数据、更多技术细节以及具体实施方法的论文正在撰写中。我们认为，在这类实验得出结果之前，先阐明这些联系十分重要，这样科学界和热心的普通读者都能轻松了解这项工作的思想。这是迈向技术未来的一小步，却令人振奋——一种以 SOULS 打造的技术。

目的论、TAME 与 SOULS

任何存在时间超过 1 个普朗克时间的“事物”，都必定在“做某件事”，使其能够持续存在。这就是“自由能原理”（FEP）最一般的形式。自由能原理最初由卡尔·弗里斯顿提出，用于解释大脑的行为，之后其适用范围扩展到了有生命和无生命的“事物”。用数学形式来说，我们称一个“事物”为一个“动力系统”，具体而言，是处于“非平衡稳态”（NESS）的动力系统；而系统“所做的事”使其能够维持在该状态 [12]。系统“所做的”这件“事”通常可以称为“自我证据化”。卡尔·古斯塔夫·亨佩尔在讨论对事件的解释时引入了这个术语。顾名思义，它是某种事物“自明”的动词形式。如果某个解释所解释的事件为该解释提供了关键证据，那么这个解释就是自我证据化的 [54, 22]。事实证明，我们所能观察到的每个“事物”之所以能够持续存在，其解释中很大一部分可以

必须根据其所做之事来理解。简而言之，如果某个“事物”没有在一定程度上参与这一自我证据化过程，那么它就无法持续存在。

同一脉络下的一个相关论点是：某个事物若要存在——即使它的存在并不持续——它也必然必须具有因果力。换言之，它必须对其环境施加某种影响。这被称为“埃利亚原则”，最早由柏拉图在其著作《智者篇》中描述 [40]。自由能原理只是将这一点延伸为：当考察任何能够**持续**其存在的事物时，人们会发现，经测量，它所施加的因果力平均而言必然是自我证据化的。



- “自我证据化的因果力”的简化示意图。环境波动导致边界内的粒子“过多”，从而使其中一个受约束的粒子被排出。
- 描述中的一个关键短语是“偏置环境”。结构未必能够保证自身的创造。也可能只是结构提高了环境在一段时间内重新创造该结构的概率。

通过在某个“事物”的特定动力学与变分推断之间建立形式联系，自由能原理使我们能够以目的论的方式讨论通过某种自组织过程创造的“事物” [15]。也就是说，讨论相对于其“目标”的“事物”。“目标”一词可以从严格的控制论意义上理解。在一个控制论系统中，系统的目标是调节某个特定变量的值，就像恒温器调节温度一样。系统的调节行为被视为“目标导向的”。由于一个“事物”为了维持自身存在，必须调节的最基本变量就是其自身的特定动力学，因此我们确实可以说，它的行为是目标导向的。因此，自由能原理表明，每个“事物”的首要目标都是自我显现、调节其特定动力学，或更具体地说，通过变分推断过程最小化其变分自由能。关于如何建立上述形式联系的数学细节，不在本文讨论范围之内。读者可参阅 [15] 以获取更多信息。

构建“事物”的特定动力学与变分推断之间这种形式联系的重要意义在于，它为我们更好地理解任何“事物”的行为提供了指导原则；这种“事物”的智能与认知取决于其执行此类自我证据化变分推断的能力。换言之，所有“事物”都表现出某种程度的智能与认知。事实上，任何万有理论中的基本单元，即粒子，都必然具有这种“通过推断进行自我证据化”的特征，才能使其存在持续于时间之中，因此，至少在最基本的意义上，它们也具有一定程度的智能与认知。

这里有几个术语值得进一步讨论，即智能和认知的定义，以及本文为何倾向于目的论而非目的拟态的论证。智能和认知这两个术语的定义一直充满争议，其情形与“意识”一词的含义所经历的争论相似。因此，本文不对这两个术语给出一个所有人都认同的严格定义。相反，本文采取一种操作性立场，根据支撑自由能原理的变分推断数学来定义它们。在描述任何“事物”时，自由能原理的语言会谈到它的“信念”；这些“信念”本质上构成了该“事物”关于其“感觉知觉”成因的知识。本文进一步指出，任何

“事物”都必须“做某件事”，也就是基于其“信念”对外部环境采取“行动”。在本文中，将“感觉知觉”和“信念”转化为针对外部环境的“行动”，且这些行动平均而言服务于自我证据化的过程，就是认知。于是，可以通过该“事物”能够“感知”“相信”和“行动”的程度来衡量认知。换言之，可以分别通过回答如下问题来衡量认知：“它感知到什么，以及感知的细节程度如何？它似乎持有什么信念？”以及“它有哪些可采取的行动？”那么，智能就是这种“感知、相信、行动”循环的一种特质。具体而言，智能是某个“事物”实现其存在“目标”的适应能力。换言之，如果我们看到某种存在之物，那么它越接近以某种特定方式走向不存在（“死亡”），即它所承受的“应力”越大，并且越能将自身拉回“健康状态”，我们就越会认为该事物具有智能。某个“事物”能够从“应力状态”回到“健康状态”的能力，必然意味着该“事物”能够更新其“信念”，从而改变其行动。这是因为它必须做出不同于导致其进入“应力状态”的行为。因此，当该“事物”在某种程度上更善于从特定类型的“应力状态”中将自身拉回“健康状态”时，我们就会说它已经学会了。一个“事物”可能遭受“应力”的各种方式，体现了一个系统展现智能的各种机会；因此，智能是认知的一种多维特征。

“目的拟态”一词旨在区分拥有“真正目标”的“事物”（目的论系统），如人类，和仅仅看起来“仿佛”拥有“真正目标”的“事物”（目的拟态系统），如机器人或细菌。本文认为，这种区分并无必要，依据是自由能原理在“事物”特定的动力学与推断之间建立的形式联系、对“目标”这一控制论含义的运用，以及以下聚焦于所有事物共同内在特征——存在——的论证。

关于根本目的论的论证:

1. “目标”是某种事物的行为所期望实现的结果。 2. 根据某物的行为来识别其目标，是一个尚未解决的问题，在某些情况下甚至可能是一个难以处理的问题 [**5**, **2**, **24**]。 3. 然而，任何已经存在的“事物”，为了能够存在，都必然表现出使其得以存在的行为。 4. 因此，存在是任何已存在的“事物”之行为所产生的结果。 5. 那么，问题就是：这个“事物”是否“意图”存在？存在是否就是它“期望实现”的目标？——我们希望知道情况是否如此，因为我们只能通过观察某种行为相对于某个目标的表现来衡量智能。只有两种可能的答案：a. 存在就是它的目标，在这种情况下，观察其行为便能使我们衡量它的智能。b. 存在并非它的目标，在这种情况下，观察其行为便能使我们衡量它的**无能**。

虽然对第 5 点提出的“意向性存在”问题所给出的两种答案，就其对测量某个“事物”的行为所允许的范围而言是等价的，但二者所允许的解释不同。只有第一种答案的解释与我们通常理解的“智能”恰当地一致。

读者可能已经注意到，关于“事物”的存在目标，这一定义的智能存在一个潜在问题。如果你观察某个“事物”，而它死了，这是否意味着它必然在某种程度上能力不足？这是否进一步表明，自愿赴死的“事物”能力不足？例如，一位为拯救队友而牺牲自己的战争英雄，又该如何看待？答案是，人们必须拓宽观察视角。本文对智能的定义要求人们恰当地识别研究对象这一“事物”；对于舍身牺牲的战争英雄而言，脱离其所处的背景，便无法恰当地理解这类牺牲事件。本文建议，不要将战争英雄视作一个独立的“事物”；在观察到此类牺牲事件时，必须转而结合其所属团队或集体的背景来考察，而此时，具有更高层级智能的团队或集体便成为研究对象这一“事物”。在这一背景下，战争英雄不一定会被视为“能力不足”，因为他们是作为更大组织的一部分在行动，而这个组织可能才是恰当的研究对象这一“事物”。我们也可以将有机体的保护性皮肤细胞与有机体本身作类比。皮肤细胞的牺牲行为并不能成为判定其能力不足的依据。

这种基于存在论的多尺度分析对智能意味着什么，会影响我们如何理解某个特定的个体“事物”或“自我”。

当构成任何“事物”的各个组成部分朝着相同的宏观层面的目标状态运作时，自我便形成了 [**10, 49, 1**]。最重要的是，这些组成部分只有遵循最小作用量原理，通过自身自然的物理行为，才会协同运作 [**32**]。建立连贯的“自我”的一个重要方面，是共同分担应力。分担应力会使背叛变得不可能，从而促成合作而非竞争。例如，如果各组成部分在应力方面彼此隔离，那么组成部分 A 就可能使组成部分 B 承受应力，而自身的压力水平不会增加。然而，如果各组成部分通过物理力或分子共享应力，那么组成部分 A 使任何其他组成部分承受应力时，自身的压力水平也会随之增加。组成部分之间共享应力的程度越高，背叛就越不可能发生，“自我”也就越连贯。通过应力将各个独立的“自我”联结起来，可以使更多资源因此投入同一目标，从而提升计算能力 [**30, 29, 11**]。这或许可以用下面这句话来概括：“令你承受压力之物，也令我承受压力，我们便是一个自我”。

正如我们从前面的讨论中所看到的，“智能”“认知”“目标”“应力”和“自我”等概念，在从基本粒子到人类乃至更广范围的所有空间尺度上，对所有“事物”都保持不变。此外，这些概念的定义完全符合自然主义，并且具有可操作性。换言之，它们不仅是有助于测量和执行工程任务的实用概念，也不需要任何超出最基本层面物理定律所包含内容的东西。从这个意义上说，自宇宙诞生以来，宇宙的演化就是一个通过各种“事物”的组合，将这些基本属性逐步扩展的过程。

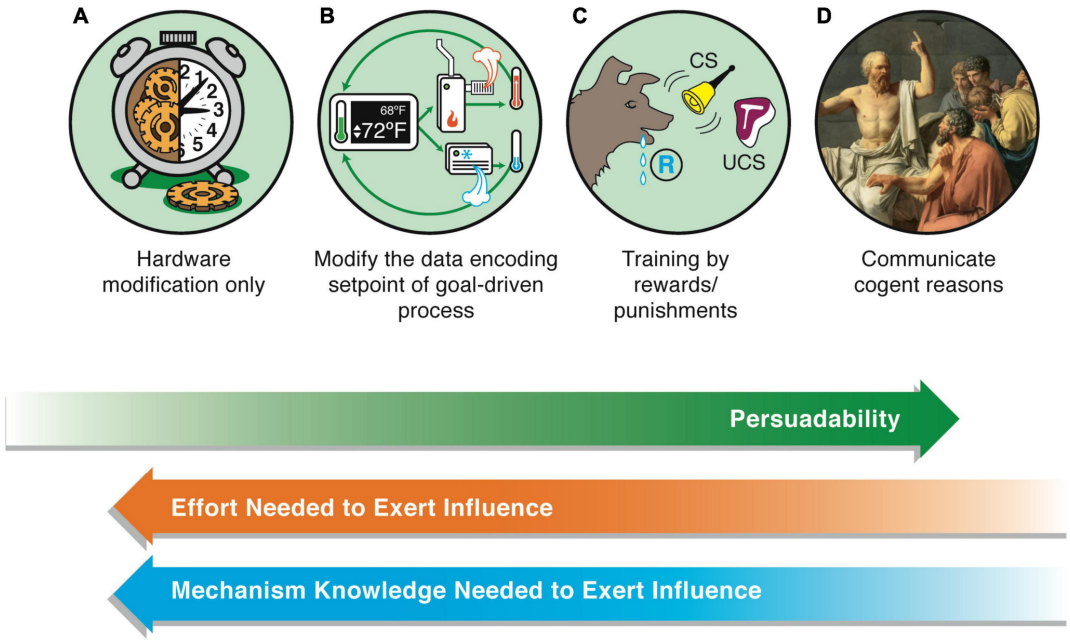
更复杂的“事物”产生的过程，可以理解为一种广义进化：在这一过程中，自然选择的对象是所有“事物”，而不只是“生物体”。在这种广义进化中，我们谈论的是“事物”的稳定性，而非“事物”的适应度。与其说“只有最适者才能生存”，不如说“只有最稳定者才能延续” [**53**]。换言之，稳定性更高的“事物”会“胜过”稳定性较低的“事物”。组装理论很好地描述了这一广义进化过程；在组装理论中，这一过程称为“组装过程”，我们将在“开放式进化、组装理论与建模”一节中详细讨论 [**36**]。

鉴于：（1）“事物”的组成部分通过自身的行为协同运作；（2）这类“事物”及其组成部分都可以被描述为具有“智能”；（3）所有事物的行为都只遵循最小作用量原理，因此，组装过程中形成的任何“事物”似乎都可以被最恰当地描述为**自组织**，因为其结构源于原本相互独立的部分之间的局部互动；就其满足存在目标的适应能力而言，它是**智能的**；就其实现同一存在目标所需的能量而言，它是**超低功耗的**。因此，这类结构最适合被称为自组织智能超低功耗系统，即 SOULS。

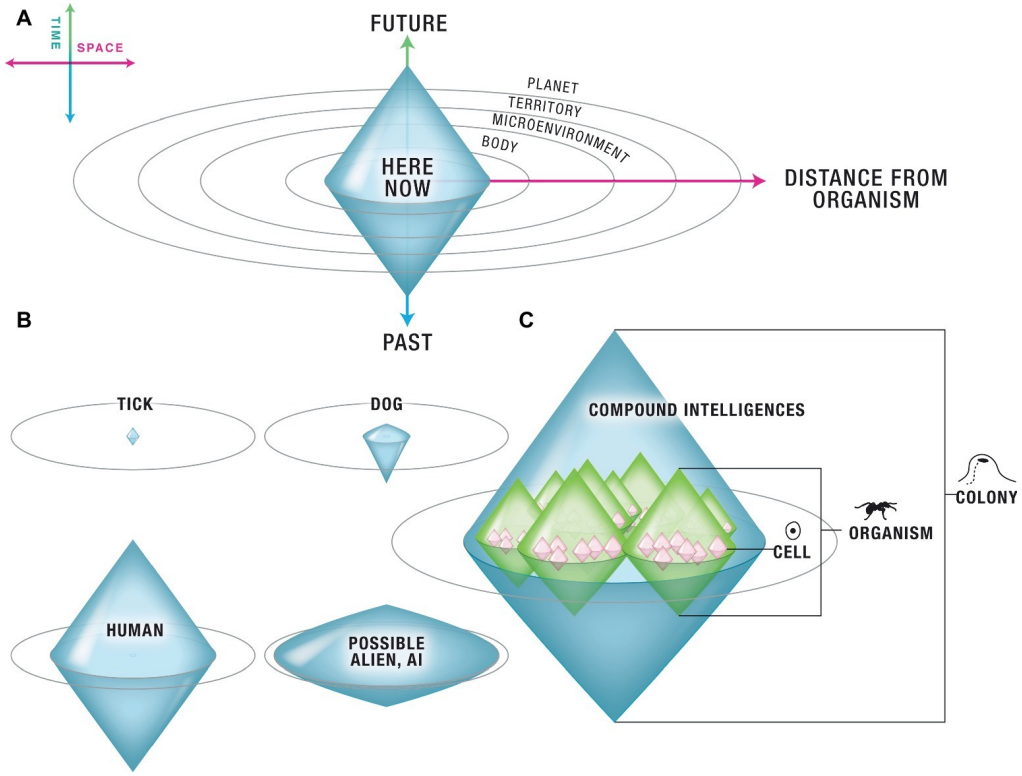
本文关注的是，作为人类，我们能够如何协助“事物”相互组合，使我们构建的成果成为可用于科学和工程领域的实用技术。如前所述，所有“事物”固有的不变多尺度特征打破了离散类别的界限，将“事物”置于一个连续谱上。在这个连续谱中，实验者或工程师必须思考：“我所研究或用于工程设计的这个‘事物’，具有多少、哪种属性 [**30, 49, 29**]？”因此，用来指称相关“事物”的术语，最能反映观察者的目标，而非“事物”本身的固有属性。在工程领域，“事物”可以称为“结构”“材料”或“组件”；在生物学领域，可以称为“化学物质”“有机体”或“生命”；而正如我们将在 Ensoul 的案例中看到的，它们也可以称为能量稳态循环、网络和巢，这些都是不同尺度下 SOULS 的抽象表述。

普遍心智的技术路径（TAME）是由 Michael Levin 博士正在开发的一个框架，旨在为理解多样的身体与心智，或 SOULS，提供一个以实验为依据的框架。该框架的目标是利用此类不变量，作为一种思考方式，探讨我们如何创造并以之为基础进行工程设计，构建具有不同程度智能和认知的结构。TAME 中一个重要的概念是“可说服性轴”。Levin 表示：“可说服性指的是能够理性地改变特定系统行为的最佳概念与实践工具。”有些 SOULS 比其他 SOULS 更容易被说服，而其可说服性在一定程度上取决于 Levin 所说的 SOULS 的“认知光锥” [**30, 49, 29**]。认知光锥代表 SOULS 的智能或解决问题的能力所能纳入考量的时空范围。例如，简单的单细胞生物

可能只会“在意”局部糖浓度梯度之类的事，以及发生在毫秒乃至几秒时间尺度上的事件。相比之下，人类会在意数十亿年后宇宙将会发生什么（时间跨度很长、空间范围也很大）。因此，人类的认知光锥远大于简单单细胞生物的认知光锥。



- “图表改编自 Levin（2022年）发表于 Frontiers in Systems Neuroscience 的文章，依据知识共享署名许可协议（CC BY）发布。来源：https://www.frontiersin.org/articles/10.3389/fnsys.2 022.768201/full”



- "图改编自 Levin（2019年）发表于 Frontiers in Psychology 的文章，依据知识共享署名许可协议（CC BY）发布。来源：https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02688/ full"

出于实际考虑，按每个目标分别定义可说服性轴会很有帮助。根据目标的不同，某些 SOULS 可能比其他 SOULS 更合适，而这与它们自身的可说服性无关。例如，如果目标是让 SOULS 烘焙饼干，那么一台典型的计算机就需要被拆解成各个组成部分，并可能重新配置成一台小型制饼机器人。这在可说服性轴上的位置会较低，因为它需要进行强制性的细致管理。然而，只需告诉一名真正的厨师要制作什么，也许再用一些钱说服他即可。由此，对于某个目标而言，可说服性轴与实现该目标所需的信息量呈反向关系。换言之，TAMEer 为使 SOULS 实现某个目标而必须耗费的能量，会随着 SOULS 在可说服性轴上的位置向右移动而减少。



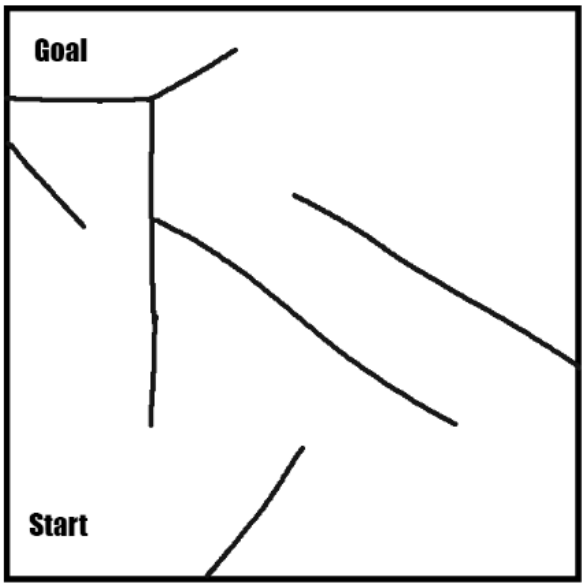
- 可说服性与能量之间的关系：需要投入多少以知识/努力形式存在的能量，才能使 SOULS 与你的目标保持一致。
- “图表改编自 Levin（2022 年）发表在 Frontiers in Systems Neuroscience 上的论文，依据知识共享署名许可协议（CC BY）发布。来源：https://www.frontiersin.org/articles/10.3389/fnsys.2 022.768201/full”

重要的是要认识到，可说服性轴是一个连续谱。我们讨论了该轴最左端的微观管理，以及最右端的传达有说服力的理由，但处于中间位置的 SOULS 也同样存在。可说服性轴图示中列举了两个例子：恒温器和狗。如果一个人的目标是改变所在房间的温度，那么恒温器就是一个不错的选择；为此，你不需要对它的所有组件进行微观管理，只需改变设定点即可。同样，如果一个人的目标是让狗从冰箱里给自己拿一瓶啤酒，那么对狗的所有肌肉进行微观管理，或试图通过脑机接口控制其神经元，可能都不是最佳方法。相反，人们可以用零食（奖励）训练狗。在这里，智能确实取决于观察者的判断。人们必须认识到，自己的目标与所互动的 SOULS 的目标是如何保持一致的。这种一致性越高，就越容易“说服”SOULS 去做你希望它们做的事。有关 TAME 以及其他用于识别特定 SOULS 更高层次“非存在性”目标的工具，以便对其进行修改并与之协作的内容，将在“驯服 SOULS”一节中进一步讨论。现在，我们先转向其他与发展能静态网络相关的概念，将其作为创建 SOULS 的工具。

开放式演化与组装理论

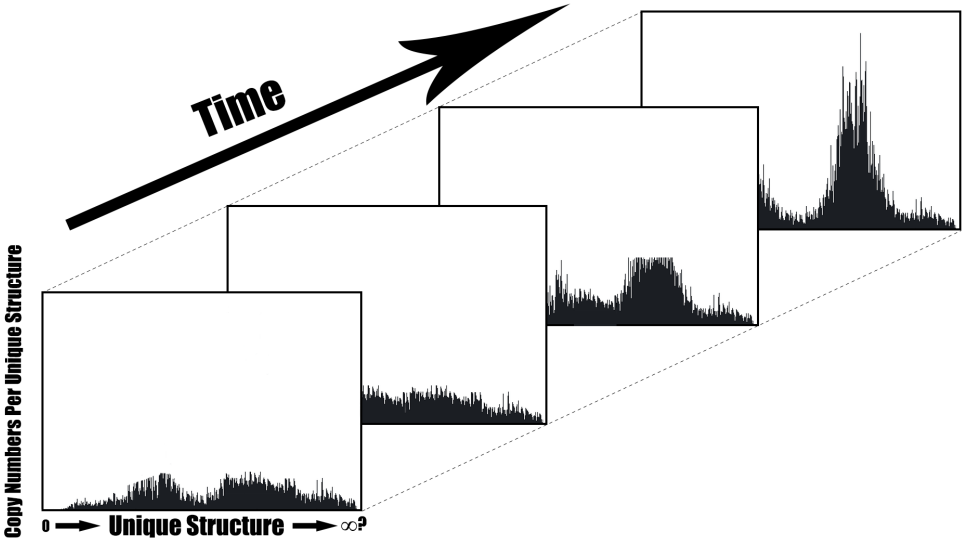
尽管标题如此，进化编程领域至今仍无法捕捉到我们环顾宇宙时所观察到的真正的开放式进化。Kenneth Stanley博士认为，原因在于大多数进化算法仍然专注于优化而非发散，而优化的问题在于，最宏大的目标在实现路径上具有欺骗性 [28]。换句话说，如果你的目标是创造类似通用人工智能（AGI）的事物，那么通往目标的道路绝非一帆风顺。实际上，会有许多看似具有欺骗性的子目标，表面上似乎能让你更接近目标，实际上却会让你离目标更远。训练神经网络执行某项任务时，我们也能看到这种现象的端倪。神经网络可能会陷入局部极小值，并“相信”自己正朝着正确的方向前进，但实际上，这只会让它离全局极小值更远。一个很有实践意义的欺骗性示例，是训练双足机器人行走。你可以采用目标优化的方法进行训练，以逐步接近行走为目标；也可以采用发散探索的方法。Stanley博士运用了这两种方法，发现通过发散探索“训练”的双足机器人，最终比专门针对行走进行优化的双足机器人走得更远 [28]。一个更直观的例子是训练

神经网络解决迷宫问题。如果你只是通过优化来缩短智能体与目标之间的距离，那么下面这个迷宫就构成了一个欺骗性示例：



技术的大部分发展都聚焦于构建具有特定期望输出的机器或程序。也就是说，技术的创造主要聚焦于收敛，而非发散。正是这种对收敛的关注，使得当今大多数技术成为“生命证据”，而非“生命”本身。在组装理论中，“生命证据”指组装指数高于某个阈值的结构。组装指数是衡量结构复杂性的内在度量 [**35, 48, 37**]。其理念是，与通过有偏过程相比，高复杂性结构通过随机无偏过程产生的可能性要小得多，因此构成了某种稳定构造器存在的证据。这种稳定构造器会被视为“生命证据”，也确实是一种生命结构；或者换言之，是一种位于认知连续谱相对靠右位置的 SOULS。SOULS 能够构造的结构越复杂，就越可以说它越“有生命”。由此，“有生命”这一概念被置于一个连续谱上。

从更宏观的角度来看，在组装理论的语境下，生命本身可以被描述为这样一个过程：生命证据以可能不同的复杂性层级生成。从某种意义上说，我们的宇宙是“有生命的”。由于组装理论是一种物理理论，我们可以设想其他可能的宇宙，它们可能比我们的宇宙更“有生命”或更“不具生命性”。可以说，一个宇宙的“生命性”、“有趣性”或“智能”水平，取决于该宇宙穿越“组装空间”的轨迹。具体而言，最具生命性、最有趣且最智能的宇宙，应当是其轨迹平均而言最能最大化每个可能复杂性层级上的独特对象数量的宇宙。有趣的是，根据 FEPs 对物理相互作用之信息对称性的描述，任何结构的环境（即宇宙）本身都必须被视为一个智能体（我们将其称为“U”）。此外，随着 U 获得其中各种结构的副本，它便能更好地预测这些结构的行为，从而更好地最小化其变分自由能（VFE） [**13**]。如果确实认为智能是智能体最小化其 VFE、从而维持自身存在的能力，那么最智能的宇宙就是这样一个宇宙：其中每个已存在的独特对象的副本数量都得到最大化。对于如何在组装系统中更严格地测量上述性质，超出了本文的讨论范围。



• 展示智能宇宙穿越组装空间的轨迹可能呈现的样貌。横轴表示与每个独特结构相关联的索引。纵轴表示与每个独特结构相关联的拷贝数，即当前存在多少个该结构。每一帧代表一个不同的时间点。每个结构—索引对均按照如下方式排序：右侧的结构比左侧的结构具有更高的复杂性。这形成了某种“波”，其中结构推动了更复杂结构的生成，同时也体现了资源约束。换言之，在资源约束下，更复杂的结构会排斥大量低复杂性结构的存在（因此形成了这道波）。然而，在没有资源约束的情况下，这道波原则上可以永远传播下去；因此，图表右下方出现了无穷符号—问号组合。还应注意的是，这种无限传播并不必然要求低复杂性结构减少，从而或许会提高系统的“有趣性”和“生命性”；其中，有趣性指现存结构总体上的多样性与数量，而生命性特指生命结构或稳定构造器的多样性与数量。

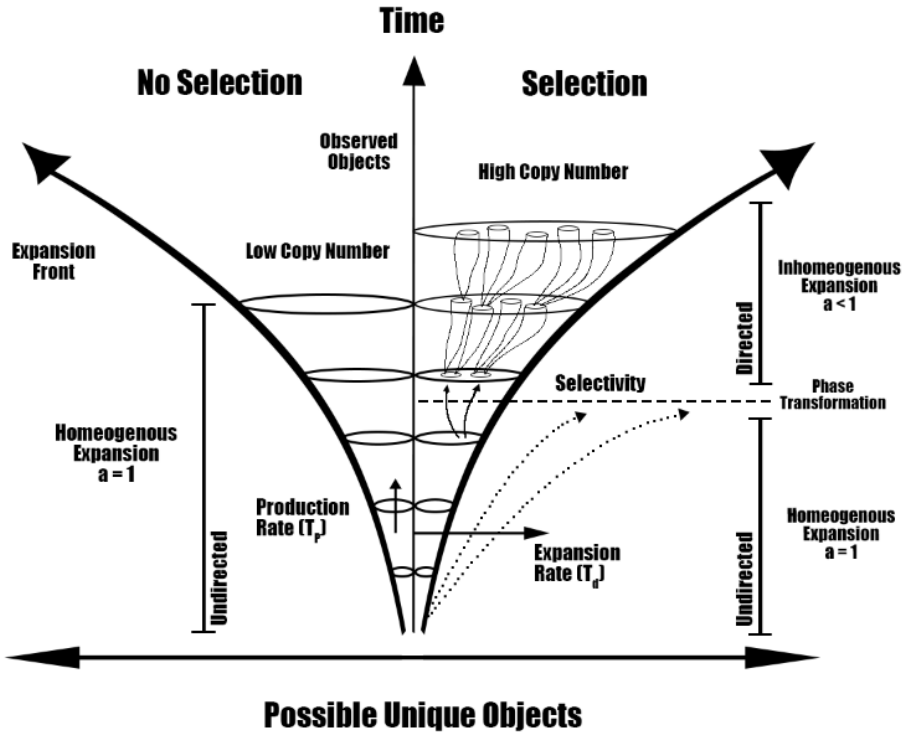
因此，生命本身就是 SOULS 组装并重新组合、创造出复杂程度各异的更多 SOULS 的过程（这些生命证据本身也可能具有生命）。这里的建议是，我们应该更多地关注如何创造能够创造更多技术的技术。换言之，如何创造智能、开放式演化的生命过程。这将帮助我们避免欺骗性子目标、发现新技术，并通过研究此类过程所创造的结构，进一步了解现有结构，也就是 SOULS。

值得注意的是，组装系统本身就体现出**自组织**，这体现在其内部动力学所产生的后果上；它也体现出**智能**，这体现在其创造并维持高复杂性物体的能力上；还体现出**超低功耗**，因为其动力学的“运行”（即其组件施加因果力）总共需要 0 能量 [**52**]。因此，组装系统本身可以被视为一个自组织智能超低功耗系统，或 SOULS。

通过生物学实验和观察，人们已经强烈暗示了开放式进化的要求 [**49**, **28**], , 而组装理论 [**48**]则从现象学角度对其进行了很好的描述。进化并不只是采用随机突变，还会通过包括但不限于增强的表观遗传能力、转座、水平基因转移和杂交等机制，采用偏向性突变。这些机制使进化产生的生物体能够通过日益智能化的突变操作，实质性地参与自身的进化。这一相互的偏向性过程可以由组装理论很好地描述。

根据组装理论，通过考察组装新颖对象的发现过程，即在一段时间内发现新对象的速率（发现时间尺度 - t_d）（k_d），以及在一段时间内生产特定对象的速率（生产时间尺度 - t_P）（k_P），可以获得有关所谓“组装空间”的信息。当比值 k_d/k_P 远大于一时，这会导致独特对象数量激增，但副本数较低（拷贝数较低）；而当 k_d/k_P 远高于

一方面，我们观察到一些具有高拷贝数的常见对象数量众多。任一极端都会导致缺乏能够生成更复杂对象的轨迹 [48]。可以证明，物理系统中开放式进化的出现，可以通过从 $k_p < k_d$ to $k_d < k_p$ 的转变来检测，换句话说，就是从对接下来会创建什么不带偏向，转变为对接下来会创建什么越来越有偏向。



- 展示选择作用下从无偏复杂性增长到有偏复杂性增长转变的图示。该图根据《组装理论解释并量化选择与进化的出现》重新绘制，来源：https://arxiv.org/ftp/arxiv/papers/2206/2206.02279.pdf [48]

本质上，组装理论认为，开放式进化算法遵循以下形式：通过无偏探索生成独特的低层级构件，而这些构件可能会限制其他低层级构件出现的可能性；随后，这些低层级构件组合成“更高层级的构件”，其中一些构件能够通过其内在的因果动力学跨越时间维持自身。由此引发的两个问题是：

1. 这些低层级构件是什么?
2. 是什么促成了这些构件组合并被选择，进而形成更高层级的构件?

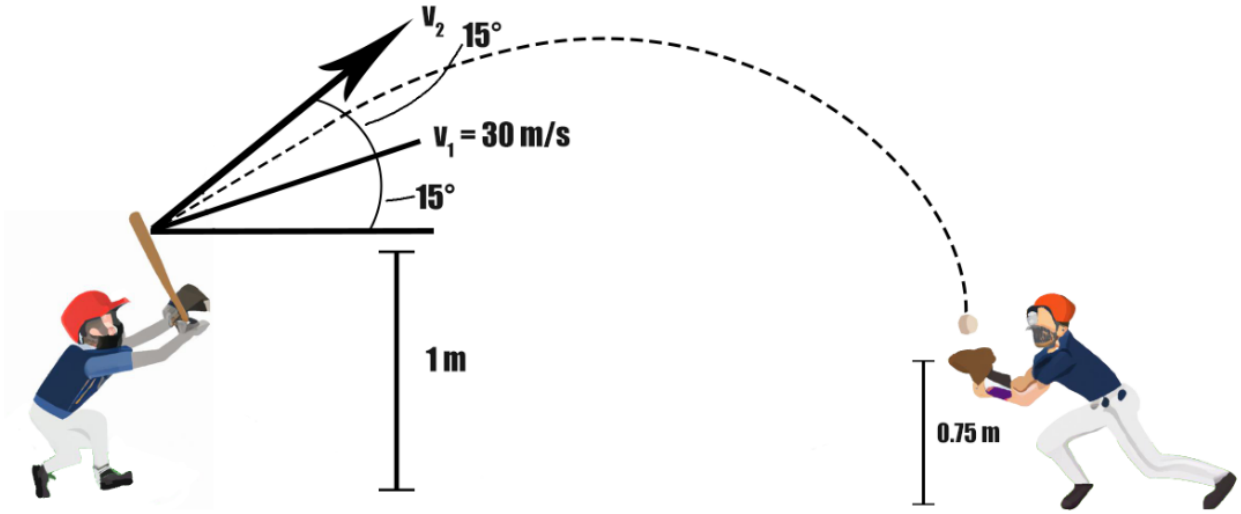
一个有助于理解的类比是：将低层构件视为类似于编程语言中的原语，而将负责组合和筛选这些构件的实体视为受编程语言语法或规则约束的软件工程师。软件工程师将这些原语组合成越来越复杂的函数，再根据其效用进行筛选，并用它们构建越来越复杂的软件。重要的是，所使用的编程语言会不断演化和多样化，软件工程师据以运作的思想和模型也同样如此。实际上，这种“类比”只是一个更大型组装过程的子集，而这个过程就是宇宙演化。本文着重探讨如何利用能量稳态循环，对一种能够呈现“可能的”宇宙演化的组装过程进行建模。不过，我们首先会说明为何此类循环非常适合作为低层构件，解释其运作方式，并展示一个应用案例。

IOS Illusions: Motivating Modeling With Enerstatic Loops

一个后果是，所有存在的“事物”都必须“做些什么”才能存在（即进行自我证据化），因此，大多数“事物”并不能像大多数典型计算机那样，最好地被描述为“输入/输出系统（IOS）”。相反，它们更适合被描述为SOULS：它们调节着各自特有的动力学，并在此过程中对环境施加因果力，而这种因果力看起来就像是它们针对某些输入所产生的输出。

IOS 与 SOULS 的区别，最能通过它们各自解决所面临问题的方式来说明。IOS 通过解析解解决问题，而 SOULS 则通过非解析解解决问题。针对任何给定问题的解析解，都意味着对该问题进行显式建模，使问题中每个“重要”部分在模型中都有相应部分。在这种情况下，相关的“事物”拥有一个针对待解决问题的显式模型。非解析解则意味着“事物”与其问题之间存在合乎规律的对应关系，因此，构成该“事物”的各部分凭借其自然动力学，无须显式表征问题便能解决问题。在这种情况下，该“事物”本身就是待解决问题的模型。这两种问题求解方法也分别被称为“弱预期”和“强预期” [51]。

这两种截然不同的问题求解方法有一个很好的例子：它们在“外场手问题”中的应用。该问题问道：“外场手如何接住高飞球？”



- 使用“分析”解决方案阐释“外场手问题”：测量棒球的初始条件，并利用抛体运动方程计算其轨迹。

针对这一问题，分析方法可能包括使用表征性方程创建数学模型。例如，某个“事物”可以估算棒球运动员击球时的力量和角度，并将这些数值代入表示抛体运动的方程中。在这种情形下，这个“事物”必须用相应的物理部件分别表示上述方程中的每个输入变量以及方程本身。这些部件的动力学将构成对输出的计算，而该输出随后必须用于驱动行为。

针对外场手问题，非分析方法可以是选择棒球抛体运动的某个特定方面，并控制该变量，使其测量值保持不变，从而达到期望的结果。例如，某个“事物”可以观察球在空中飞行，并试图让球在其视觉范围内保持恒定的速度，使球看起来既不会加速也不会减速。其原理是：如果球正在加速，它就会落在这个“事物”的后方；如果球正在减速，它就会落在这个“事物”的前方。在这种情形下，这个“事物”不需要显式表征；它只需要能够测量速度的变化，并利用这种变化作为反馈来驱动行为，使测得的变化减小。这一提出的解决方案称为“光学加速度抵消”方法，最早由 Seville Chapman 提出 [7]。

在理解分析方法与非分析方法时，这一区分至关重要

和建模相关的“事物”时，分析方法与非分析方法之间的区别至关重要。它将关注点从“这个事物正在实现哪些算法？”转移到“这个事物正在控制哪些变量？”。事实上，将 SOULS 当作 IOS 来处理是一种范畴错误，可能导致挫败感，因为这会让人误以为“事物”的行为高度易变，要么是由于其输入中存在噪声，要么是由于内在的非确定性 [**32**]。自然系统采用非分析方法而非分析方法所获得的益处是显而易见的。非分析方法更简单、更快速、适应性更强且能量消耗更低，同时仍能受益于预测所提供的大部分能力 [**51**]。它们通过避免明确表示所要解决问题的各个部分，并避免耗费能量去计算和重新计算推测性的分析解来实现这一点。这表明，在自然界中，SOULS 很可能比 IOS 更为普遍[**43**]。

事实上，虽然 IOS 在自然界 [**7**] 和技术领域中都确实存在，但构成这些 IOS 的部分是由一组 SOULS 组成的，它们的“高度可变行为”被尽可能地压制，从而实际上能够表现得像一个 IOS。在计算机工程中，这种“高度可变行为”被称为“噪声”，并有许多不同的形式，例如热噪声、闪烁噪声、散粒噪声和量子噪声。抑制这类噪声是构建快速、可靠计算机的重要环节，因此，人们会采用滤波、屏蔽、接地、增加噪声容限、纠错码和时钟同步等技术来实现这一目标。事实上，抑制由热波动导致的噪声如此重要，以至于它被列为可能导致技术进步不再以指数速度加速（即所谓摩尔定律）的一项首要因素 [**26, 23**]。

本文提出，无论一个“事物”是生物学的有机体，还是计算机中的电路，其“高度可变行为”或“噪声”都是范畴错误所产生的观察结果。换言之，这是因为我们把一个“事物”视为 IOS，而实际上将其视为 SOULS 更为恰当。从这个意义上说，IOS 是一种幻象，SOULS 可以像对某些无理数的“实际值” π 只能以任意精度近似表示一样，以任意程度的可靠性对其进行近似。这种现象在“知觉控制理论”（PCT）中被称为“行为幻象” [**44, 34, 33**]。这是否意味着我们应该放弃基于 IOS 的方法，转而采用基于 SOULS 的方法？本文对此持不同看法。相反，正如我们认识到“智能”“认知”“目标”“应力”“自我”和“生命”等概念处于一个连续谱之上，我们也应认识到理想化 SOULS 与理想化 IOS 之间存在一个光谱。我们可以将某物改变自身输入的程度简称为“反馈”，其中零反馈代表理想化 IOS。

这里，“理想化 SOULS”指的是能够完美调节其所控制变量的 SOULS，因此受控变量根本不可能受到扰动，因为任何可能的扰动都已经被“预测”了。这类似于预期同步的数学模型，其中系统 A 和 B 完美耦合，使得无论 B 的行为如何，A 总能预测 B [**51**]。“理想化 IOS”指的是完全确定且完全可靠的 IOS。提到“理想化”，是承认大多数系统中似乎都存在噪声，无论它们是 IOS 还是 SOULS。例如，在这个连续谱的中间，我们可以设想一种奇特的恒温器。它有一半时间控制着其所在空间的温度，另一半时间则把自己的计算转发给某个别的系统，而那个系统会做什么，谁也不知道。

这一视角表明，在尝试理解任何“事物”时，我们首先应该判断它更像是可靠的 IOS，还是更像 SOULS。与评估任何“事物”的智能水平一样，进行这一判断时所采用的分析尺度，以及将观察到的感觉数据谨慎地划分为不同“事物”，对于正确理解都至关重要。无论 IOS 多么可靠，如果过于细致地观察它，就会发现类似 SOULS 的行为；换句话说，人们会观察到一组正在调节自身动力学的“事物”——这是分析尺度选择不当所致。同样，只关注 SOULS 的特定部分，可能会得出结论，认为它只不过是一组有噪 IOS——这是划分方式不当所致。

鉴于 SOULS 而非 IOS 似乎能够在最根本的层面上捕捉事物，并且 IOS 可以通过 SOULS 的集合任意逼近，本文提出使用一种简单的 SOULS 抽象来对任何“事物”进行建模。这种简单的抽象就是能量稳态回路，换言之，是一种主要关注管理自身“能量水平”的控制论控制系统。通过这一抽象展开探索所希望深入理解的重大问题包括：

1. 如果一切都是 由一个“事物”（能量）构成，我们如何以及为何会观察到彼此 distinct 的事物 ？

2. 将能量以不同方式划分为“事物”可能会带来哪些帮助？3. 能量稳态循环的集合如何在更高尺度上调节变量（即在更大空间尺度上可测量的变量，例如球的速度）？ 4. 我们在多大程度上能够控制并与这些集合相处?

能量稳态循环、网络与巢。

Ensoul方法并不限定能量稳态循环可以采取何种形式。关键只在于创建一个负反馈系统或控制系统，以某种方式调节自身的“能量水平”，从而对其环境施加因果力。换言之，就是创建一个能量稳态循环。能量水平是对能量稳态循环状态的部分表征，而因果力则体现了能量稳态循环与其环境之间特定的能量交换，因此使这样的循环成为一个“热力学开放系统”。我们并不关心能量稳态循环如何形成，而是观察到“自然界在能量流动之处构建事物”，并据此构建模型。至关重要的是，不要误将以下具体实现示例当成创建能量稳态循环的最终且唯一方法。可以将它们视为Ensoul框架中的思想实验，只不过这些实验恰好能够以代码形式在计算机上实现。

一个能量稳态循环可以通过能量交换与其他能量稳态循环相连接，它们共同构成能静态网络。这些网络或单个循环也可以存在于更大的能量稳态循环内部，从而构成能量稳态嵌套。这体现了现实的多尺度分形本质，但也可能使术语变得令人困惑。鉴于这种多尺度方法，我们将作为研究或建模对象的“主要”能量稳态循环，与嵌套在主要循环内部的“较低层级能量稳态循环”区分开来，并将这类嵌套循环称为“结构”。我们依据以下准则，使用能量稳态循环、网络和嵌套进行建模：

1. 存在流入系统的能量流，其流量取决于该系统的作用。2. 这种能量流会形成能量分布，使系统的不同部分以其他能量稳态循环或“结构”的形式具有高低不同的能量。能量分布的形成代表系统开始“存活”，即处于非平衡稳态。3. 这种能量流会以某种方式改变上述能量分布。4. 系统通过施加因果力产生熵，因此能量不可避免地会损失。5. 结构只能在能量密度足够高、能够支持组装的区域内组装。6. 存在一些规则，用于定义我们的能量稳态循环可组装的结构空间 a. 一个结构至少由以下元素构成 i. 组装结构的成本 ii. 拆解结构的成本 iii. 在空间中的位置 iv. 结构的因果力 v. 因果力的成本 1. 因果力始终具有成本，并且会根据该因果力在每个时间步产生的特定效应“消耗”能量。不满足这一能量要求将导致该结构不存在。

结构的因果力由以下属性构成：

b. i. 感知输入属性，这是结构“感知”到的属性集合。ii. 受影响输出属性，这是结构施加其因果力的属性集合，即被结构改变的属性。

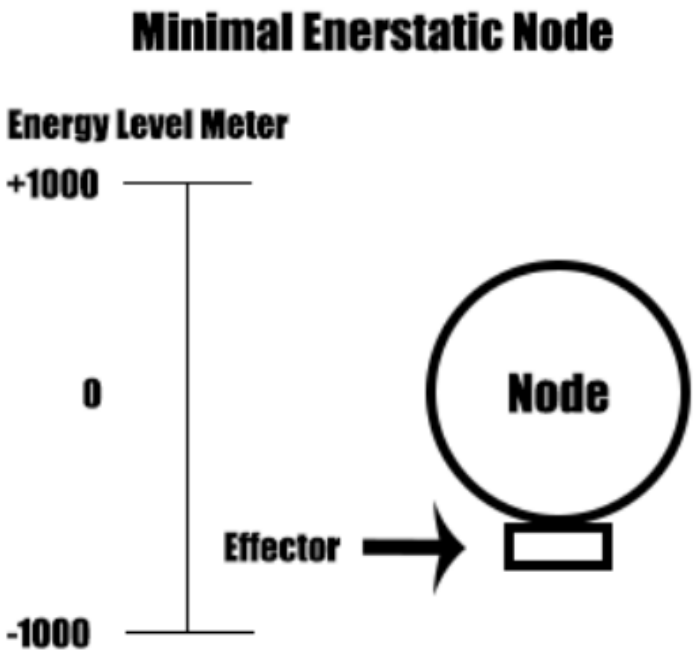
iii. 有效半径，这是能够感知或影响属性的距离。不同属性的有效半径可能不同。

当系统无法执行能够实现适当能量流动的行动时，系统就会“死亡”，即陷入平衡态。

为了更好地理解如何满足所有这些准则，我们将快速搭建一个由最小能量稳态循环构成的最小能量稳态网络。首先，我们承认，能量稳态循环是一种控制系统，形式上由四个部分组成：（1）“控制中心”，用于定义系统理想的“设定点”或状态；（2）“生理变量”，或对待调节变量的“感知”；（3）“传感器”，用于获取有关环境的信息；（4）“效应器”，用于改变环境。这让我们朝着满足此前所有准则迈出了坚实的一步。以下是实现最小能量稳态循环的可运行 JavaScript 代码，其中体现了上述所有特征：

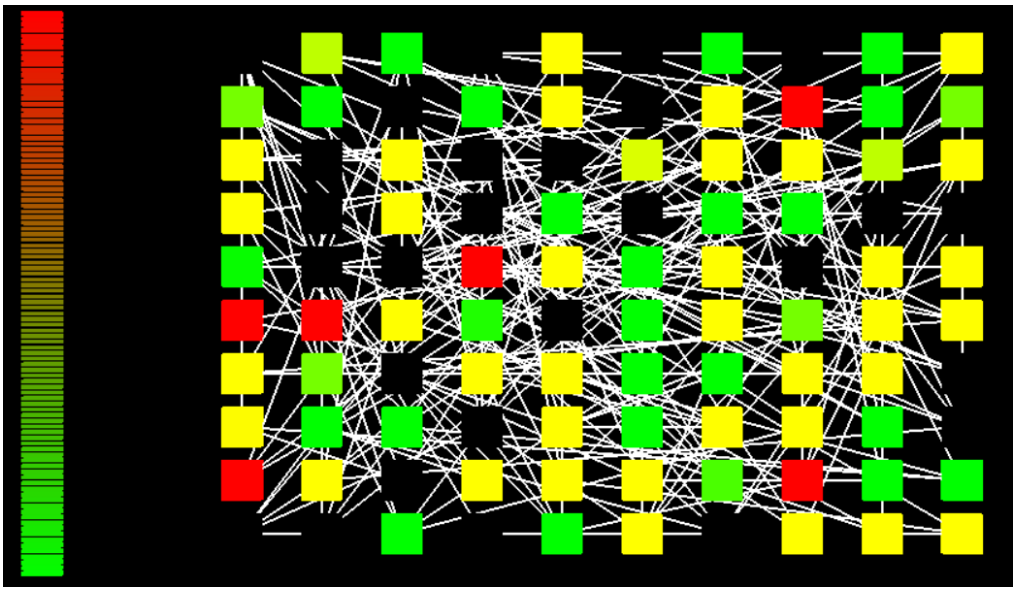
```
const timeStepToStopPerturbations = 500;
const energySetPoint = 0;
const energyConstant = 100;
const perturbationStrength = 10;
const nTrails = 1000;
const effectorCost = 10;
const effectorCount = 1;
function enerstaticNode(){
  //Energy set point defined by the "control center"
  let energyLevel = energySetPoint;
  for (let timeStep=0; timeStep<nTrails; timeStep+=1){
      console.log(`Energy at Trial ${timeStep}:`, energyLevel);
      if( isDivisibleBy(timeStep) && timeStep < timeStepToStopPerturbations ){
        //Perturbation "sensed" via causal effect on "action decision"
        energyLevel-=perturbationStrength;
      }
      //Action decision rules I.E.
      //Whether the "effector" exhibits causal power or not
      if(energyLevel > energySetPoint){
        energyLevel-=energyConstant * effectorCount; //"Effector" action
        continue;
      }
      if(energyLevel < energySetPoint){
        energyLevel+=energyConstant * effectorCount;//"Effector" action
        continue;
      }
    }
  if(energyLevel === 0) console.log("Enerstasis Reached"); //Lived
  else console.log("Failed to reach enerstasis."); //Died
}
enerstaticNode();
```

在我们的案例中，“设定点”或“控制中心”为 0，代表能量；换言之，能量稳态循环试图维持的能量水平为 0。因此，该数值既是设定点，也是需要调节的生理变量。我们的最小能量稳态循环只有一个传感器，即对其生理变量，也就是能量的感知。这是一种隐性的感知，因为当能量水平发生变化时，这会影响其拥有的“效应器”是否发挥其因果力。该循环通过环境对其生理环境产生的影响来“感知”环境，并作出反应以消除这种能量扰动的影响。



我们的最小能量 稳态循环还缺少不少能让它变得有趣且易于扩展的要素，具体包括：

7.1. 环境中只有一个循环 2. 其结构（即唯一的效应器）是固定的，无法添加或移除。
3. 这个唯一的效应器没有组装/拆卸和因果力成本 4. 未明确界定循环的死亡 5. 无法采取主动行动，也不具备任何学习能力 6. 流入系统的能量不取决于该系统的行动
7. 循环运行时不会损失能量



• 最小能量稳态网络（MEN）示例：循环健康度以循环的颜色表示。颜色越趋近红色，表示循环越接近其致命能量阈值，无论是能量过多还是过少。随着循环接近负能量阈值，其不透明度会降低，逐渐变得透明。死亡循环最终会变得不可见，但与之相关的能量通道

并非如此。每个循环都包含 3 个名为“能量通道”的结构实例，其因果力会根据起始循环的能量水平，在起始循环与目标循环之间转移能量。因此，每个循环都会随机连接到另外 3 个循环。随后可以通过各种方式扰动该网络，以观察它是否能够适应。

本文不讨论这些细节的具体代码实现，但此处提供了实现这些方面的 JavaScript 代码（用于生成上一幅图像）：https://github.com/Tyfoods/minimal-enerstatic-network

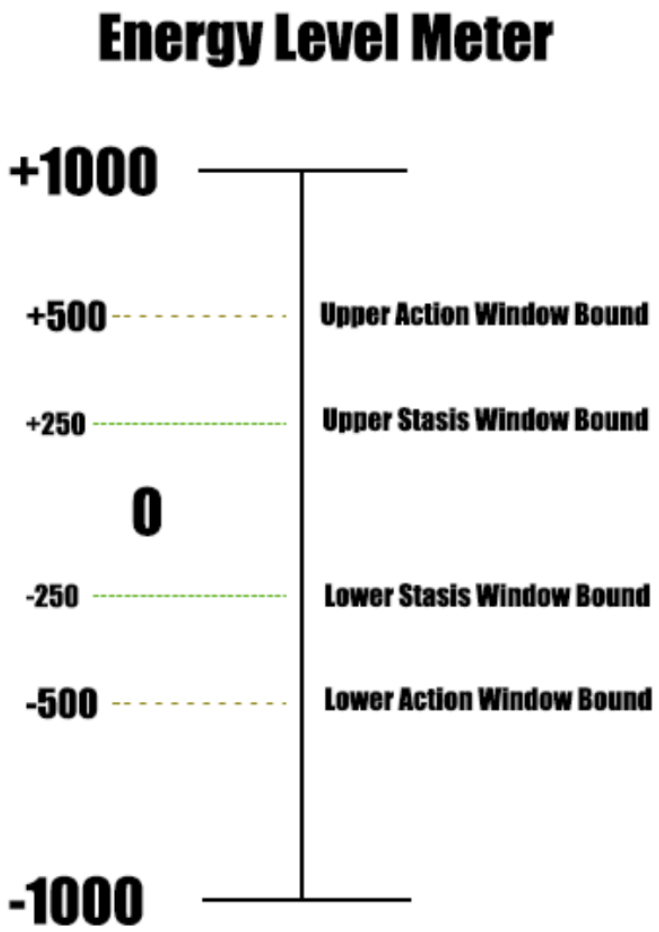
---

此时，有几条指导原则的出发点或许还需要进一步阐明。第一条是，能量稳态循环拥有一组可供其组装和拆解的结构；第二条涉及这些结构本身的性质。能量稳态循环的组织特性旨在体现组装理论中的这一观点：关于如何构造特定结构的信息，蕴含在我们所说的生命体之中。正是这种信息的具身化，使组装理论能够将生命界定为能够产生大量高于特定复杂性阈值的结构的事物；此外，它还能够将不同生命体置于“生命连续谱”的不同位置。结构是能量稳态循环的一种抽象，使人们无需涉及我们先前讨论过的各种控制系统细节，就能说明此类循环的功能。之所以可以这样做，是因为这些结构代表着始终维持在稳态设定点上的能量稳态循环，因此会持续发挥相同的因果力。我们假设，这些结构所属的能量稳态循环会将偏离稳态的结构视为“失效”，并因此将其拆解。接下来将讨论能量稳态网络中的学习算法可能是什么样的。

能静态网络中的学习

在我们的最小能量稳态网络中，每个循环都对其能量水平的任何扰动极为敏感。换言之，内部能量的任何波动都会促使循环采取“行动”。在我们的例子中，这种行动表现为单个效应器所展现的因果力。还有另一种行动形式，即对能量稳态循环进行重新配置。如果允许能量稳态循环自行重新配置，并给它时间检验新架构所带来的后果，那么我们就可以采用某种学习算法。

为 enerstatic loops 留出时间来检验其行为后果，一个很好的方法是在其允许的能量水平范围内设置两个“窗口”或区间。这两个窗口分别是“静止窗口”和“行动窗口”。静止窗口是一组围绕能稳态窗口能量设定点的数值，在此范围内，循环的架构保持不变。窗口的大小决定了循环在采取重新配置架构的行动前拥有多少时间或容差。行动窗口是一组围绕静止窗口正负边界的数值，用于指示循环应尝试重新配置其架构。



- 能稳态窗口分为两个独立的窗口：静稳窗口，在此窗口中不采取任何行动，因此回路被视为稳定；行动窗口，在此窗口中回路被视为不稳定，必须采取行动使其恢复稳定状态。

既然我们已经知道何时重新配置能量稳态循环，接下来就需要确定具体将进行哪种修改/操作、将进行多少次此类修改，以及在什么情况下进行。一种简单的方法是为每个操作定义一个概率，并规定每个时间步只执行一个操作。换言之，当循环被推入操作窗口时，每个可能的操作都对应一个概率，该概率决定该操作是否会成为该时间步所选定的操作。然而，如果循环处于操作窗口中，并不一定意味着每个时间步都必须执行某个操作。这可以由程序员自行决定，对于我们在此进行的概念性讨论并不重要。

程序员接下来必须提出的问题是：“我应该在何时以及如何更新自己的信念或概率？”在这个例子中，我们通过引入另一个窗口——“改变动作概率”（CAP）窗口，来回答这两个问题。这些窗口统称为能稳态窗口。CAP 窗口允许改变概率，作为实现稳态的最后一搏。由于目标是以一种能够恢复稳态的方式改变动作概率，设计一种试图预测应修改哪些动作的复杂算法似乎合情合理。然而，受到神经元可塑性机制的启发，我们可以选择简单的联想式赫布学习。在动作窗口期间，循环可以整合将其推入 CAP 窗口的信息或一系列步骤。具体来说，可以让循环记住采取过哪些动作，以及采取每个动作时的能量水平。我们的简单联想式赫布学习算法不会进一步探究这些数据的细节，而是根据这些动作最终使系统进入稳态窗口还是 CAP 窗口，增强或减弱在动作窗口期间采取的所有动作。这样一来，有帮助的动作概率序列会得到增强，而那些

造成负面影响的序列则会被减弱。此外，还可以记录采取动作时的具体能量水平，从而实现更精细的信念更新。

虽然探索可能的能静态网络空间本身就是一项有趣且可能具有价值的任务，但对物理现象进行建模，是深入理解现有 SOULS 并构建更加高度受自然启发的新 SOULS 的一种强大方式。例如，我们扩展能静态网络的概念，用于对神经元的功能进行建模。

神经元建模：能静态神经网络

在从概念上详细说明如何使用能静态网络对神经元进行建模之前，理解这一方法背后的生物学动机非常重要，因为它可能普遍适用于所有细胞。将细胞建模为能静态网络的目标，是探究这一理念能否作为理解细胞群体行为的指导原则。对于神经元而言，这意味着理解大脑。神经元是能静态网络发展过程中的主要灵感来源，原因在于观察到神经元在很大程度上是能够终生存续的细胞，而它们 collectively 消耗了机体能量预算的 20% [14, 20]。这引发了一个问题："既然我们看到神经元能够在有机体的一生中持续存在，那么这种神经元的行为必须是什么样的，才能做到这一点？"本文提出的答案是："在一定的稳态窗口内调节神经元能量水平的行为"。

确实，ATP 水平低到一定程度的神经元甚至无法执行细胞凋亡，而会发生坏死 [57]。相反，神经元中的 ATP 水平过高会促进细胞毒性，并最终导致细胞死亡 [16], ，或扰乱蛋白质合成，进而也可能导致细胞死亡[42]。这表明 ATP 作为"生理变量"的重要性，而作为"控制系统"的神经元正是对其进行调节的系统。

如果神经元的首要目标是控制自身的能量水平，那么可以说，神经元中的"信息处理"是"构建在"神经元调节自身能量水平所用机制之上的。在这种情况下，若试图理解神经元处理信息的方式，就会受阻于这样一个事实：神经元首先是能量管理系统，其次才是信息处理器。确实有证据表明，能量效率与功能之间存在信息性关联[39, 45]。更普遍地说，如果所有存在的细胞都必须满足这一生存必需条件，那么用能静态网络对其进行建模和理解往往是有意义的。从这一视角来看，所有细胞都不知道自己参与了哪些"计算"，也不知道自己在其所处有机体的发育和运作中扮演什么角色。细胞只是尽可能长久地存在下去，同时受其环境所限。细胞组装并维持这些结构，首要目的是让自身的能量维持在稳态范围内。这种理解上的障碍，正是《IOS Illusions: Motivating Modeling With Enerstatic Loops》中讨论的"行为幻觉"所造成的同一种障碍。现在，我们转而来看一个用能静态网络对神经元建模的例子。

由于能量静态神经网络的计算实验目前正在进行，而本节的目的只是启发人们如何用能静态网络对生物学的 SOULS 进行建模，因此只会讨论项目的一般概念细节。

能量静态神经网络本质上就是最小能量稳态网络，唯一的区别在于，能量静态神经网络中的循环所组装成的结构，会按照程序员期望的复杂性层级，模拟神经元中的结构。根据能静态网络的要求，我们可以将概念模型具体描述如下：

1. 有能量流入系统，其流量取决于该系统的活动
    - 神经元只有受到刺激时才会接收血流，也就是能量。这一设想受到神经血管耦合（NVC）的启发。

2. 这种能量流会形成能量分布，使系统的不同部分以其他能量稳态循环（即"结构"）的形式呈现出高低不一的能量。这种能量分布的形成意味着系统开始处于"活着"的状态，也就是说，处于非平衡稳态。

    - 每个神经元一开始都拥有一系列结构，例如离子通道、受体、轴突、树突、基因等.....每种结构在空间中占据不同的位置，并且创建它所需的

能量各不相同。

3. 这种能量流　动会导致前述内部能量分布以某种方式发生改变　y

- 能量由线粒体捕获和释放。尽管线粒体具有运动性，但它们在神经元周围的分布并不均匀。这会形成局部的高能量区和低能量区。

4. 系统产生熵时，能量不可避免地会损失

- 神经元具 有一种“因果力”成本，该成本被定义为它们组装的结构的函数　.

5. 结构只能在这些能量点组装

- 只有能量足够高的局部区域才能组装结构。

6. 存在定义可能结构空间的规则。

- 正如我们在最小能量稳态循环练习中所提到的，结构组装、拆解和因果力的成本由程序员决定，但也可以通过实证测量确定。同样，结构施加其因果力的条件也可以通过这两种方式中的任一种确定。因此，为了简洁起见，并且由于这与本例的重点无关，具体数值和条件均予以省略。

- **结构**

  – 宏观结构

  1. 基因 - 需要注意的是，基因组不一定会指定我们程序空间中的所有可能结构。

     - **结构基因** - 产生一种特定的“物理”结构，从而修改循环的架构。
     - **调控基因** - 规定一个基因中的基因表达（因果力施加）如何影响其他基因的表达。

  2. 胞体
     - 几何上表示为一个球体，其直径由基因指定
     - 膜电位由基因指定
  3. 轴突
     - 以二维或三维线条表示其几何形态
     - 膜电位由基因指定
  4. 树突
     - 以二维或三维线条表示其几何形态
     - 膜电位由基因指定

  – 微观结构

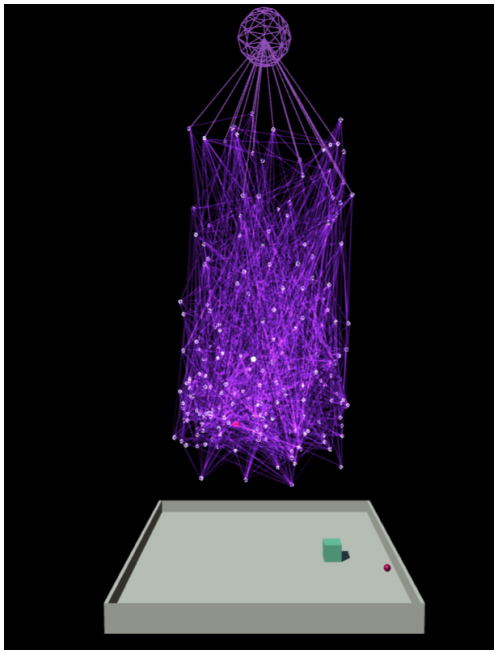
  1. 电压门控离子通道——由基因规定电压敏感性。当其嵌入的结构的膜电位达到某一阈值时激活。激活后，会使细胞的膜电位去极化（升高）或超极化（降低）；这种降低的程度由基因规定。2. 配体门控离子通道——当与基因规定的某种特定神经递质接触时激活。激活后，会使细胞的膜电位去极化（升高）或超极化（降低）。3. ATP酶泵——每个时间步长消耗一定量的能量，从而产生局部能量需求。4. 神经递质——接触后，按照基因的规定激活特定的配体门控离子通道。

5. 线粒体
   - 会根据能量梯度移动，优先前往能量较低的地方。
   - 接收到血流后获取能量，并向局部区域输出能量。
6. 信使RNA——由细胞核发送给核糖体的指令，详细说明要组装哪种结构。

7. 核糖体
   - 会根据能量梯度移动，优先前往能量较高的地方。
   - 接触后，在局部能量充足的情况下，将信使RNA翻译成一种结构 .

7. 当系统无法执行能够实现适当能量流动的行为时，系统便会死亡，即进入平衡态。

- 当神经元达到致命能量阈值时，就会发生细胞凋亡并停止运作。具体而言，它们所有相关的结构都会从程序中移除。



- 一项正在进行的实验可视化示意：能量静态神经网络接收以“稀疏分布式表示（SDR）”形式呈现的视觉输入 [3]。该能量静态神经网络必须控制球体及嵌入球体中的摄像头来完成两项任务：（1）识别食物；（2）朝食物移动——否则它就会死亡。因此，能静态网络是开放式进化的绝佳候选。

如前所述，模型参数的生物物理细节可以按需设定得尽可能详尽。这或许能帮助人们深入了解其真实生物学对应物的内部运作机制。然而，在讨论技术的创造时，人工智能的发展已表明，无须尽可能精确地设定所有参数和结构细节，也能产生既有技术实用价值、又值得科学探究的成果。事实上，尽管人工神经网络比真实的生物神经网络简单得多，人们仍利用它们来深入了解其行为。因此，利用能静态网络对物理现象建模，或许既能帮助我们理解这些现象，也能运用由此产生的技术来解决实际问题。

一种利用能量静态神经网络模型生成更具生物学相关性结果的有趣方法，或许是将抽象层相互映射。例如，霍奇金-赫胥黎神经元（HHN）比生物神经元（BN）简单，但可以利用经验数据将 HHN 拟合到 BN。同样，也可以利用 HHN 神经元生成的经验数据，将其映射到一个更加简单的神经元

模型。这样或许能使模型与实验探究更加相关。这种抽象层映射是涉及能静态神经网络的持续计算实验的一部分。

目前正在进行的其他实验，主要探索将开放式进化技术应用于一般的能静态网络。将能静态网络用作构建技术的工具，其优势在于它们能与这类技术自然共生。正如前文所述，开放式进化需要具备特定动力学特征的构件，使复杂性能够通过构建越来越大的构件而不断提升；这些构件由一种智能所选择，而这种智能本身也在系统内部构建并增强。能量稳态循环默认内置了这种选择机制，因此非常适合开放式进化。

能静态网络的进化

使能量稳态循环（进而使能静态网络）能够与开放式进化技术共生的关键特性，可概括如下：只有能够实时修改自身架构、以摄取多于消耗的能量，能量稳态循环才能维持其存在。

如果能量稳态循环获取能量的唯一途径，是根据其所处环境以特定方式成功组织自身，那么就无需为特定任务指定特定奖励，因为这类奖励本就内在于其本质。借助目前可用的工具，这使得我们难以将能静态网络设计成精确执行你所期望的任务；但这也为进化网络创造了机会，使其既能实现你期望的行为，也能展现出你可能从未设想过的行为。

开放式进化的目标是生成如同我们在宇宙中所观察到的、看似永无止境的复杂性。开发具有此类行为的系统，其挑战集中在环境—智能体困境上。这个困境在于，当嵌入某个环境中的智能体变得更加复杂，或更加智能时，它最终会掌握该环境，从而使自身的复杂性和智能达到最大值。因此，有人提出，应当创建能够与嵌入其中的智能体一同变得更加复杂的环境 [**55, 50**]。尽管这一想法看似直截了当，但如何在不定期注入某种外部智能（即程序员干预）的情况下，让环境持续与智能体一同变得更加复杂，始终难以实现。

问题的很大一部分在于环境的编码方式。尽管对于迷宫或二维景观等特定领域，已经存在通用编码，但如何将其扩展到特定环境领域之外的问题似乎依然存在。在他们的论文中，Evolution through Large Models（ELM）的作者提出：“计算机程序是一种通用且强大的编码方式，能够持续扩展现有环境的丰富性。”这一观点为展示如何利用原始计算机程序确保开放式演化无限持续迈出了关键一步 [**27**]。在论文中，他们提出了一种特定的方法，该方法能够生成一个有条件地输出计算机程序的程序，并将其称为“发明流水线”。简而言之，该方法包含三个步骤：

1. 使用任意的大语言模型（LLM）为进化算法提供智能变异（在本例中为质量多样性（QD）算法 MAP-Elites），以便在几乎不存在或完全不存在训练数据的领域中生成训练数据。2. 使用上述训练数据对 LLM 进行预训练，使其能够输出类似的数据。3. 使用强化学习以及一些额外调整，对 LLM 进行微调，使其能够根据所接触的条件生成数据。

这种方法涵盖了我们认为实现无限开放性所需的几个关键要素，具体如下：

1. 在间接编码上实现日益智能化的（相关）变异，而这些编码尽管如此，也只是共同 de.
2. 无需使用经过训练的神经网络，即可生成大量多样化的新颖结构。3. 创建一个能够根据其“需求”有条件地创建结构的智能体。

尽管形式各异，进化能静态网络旨在实现的正是这些要素。

能量封闭系统

从根本上说，能静态网络旨在表达这样一种理念：任何存在的“事物”都是一个“能量开放系统”。能量开放系统是指一种系统，它接收一定有限量的能量作为输入，并仅输出其中一部分能量。具体而言，我们的能量开放系统由以下三条规则扩展而成：

1. 必须有能量流入系统。2. 必须有能够将能量基本上困在系统内部的机制。
3. 必须有负责让能量流出系统的机制。

在这种设定下，我们面对的是一个相当静态的环境，它无法响应我们的能静态网络的复杂性而变得更加复杂。然而，我们曾提到，作为自由能原理信息对称性的结果，环境也必须被视为一个能够在某种意义上增加自身复杂性的智能体 [13]。实现这种信息对称性的一种方式，是将环境本身视为一个能量稳态循环。在这种情况下，我们现在拥有了一个“能量封闭系统”。本文指出，采用这种方法并结合 ELM 等方法，对于发展真正开放式的过程至关重要。我们的能量封闭系统遵循以下规则：

- 系统内部存在一定量的能量
- 能量既不能被创造，也不能被消灭
- 此前规定的能量开放系统规则适用于系统内的所有结构*（稍后我们会看到一个小例外，但总体上是正确的）。
- 环境被建模为一个使自由能最小化的能量稳态循环（即当前未被结构困住的能量）
- 环境能量稳态回路（EEL）负责在其中构建结构

为了更好地理解这里正在发展的方法，最好在我们所需规范的约束范围内，对我们切实可行的做法进行某种模拟或思想实验，并穿插说明我们为何可能选择以特定方式行事。当然，具体做法有很多，其中许多都可以改用于能量开放系统。遵循本文的主题：以下是一项目前仍在发展中的思想实验迭代——请将其视为“供思考的素材”，而不是一份具有决定性的“操作指南”。

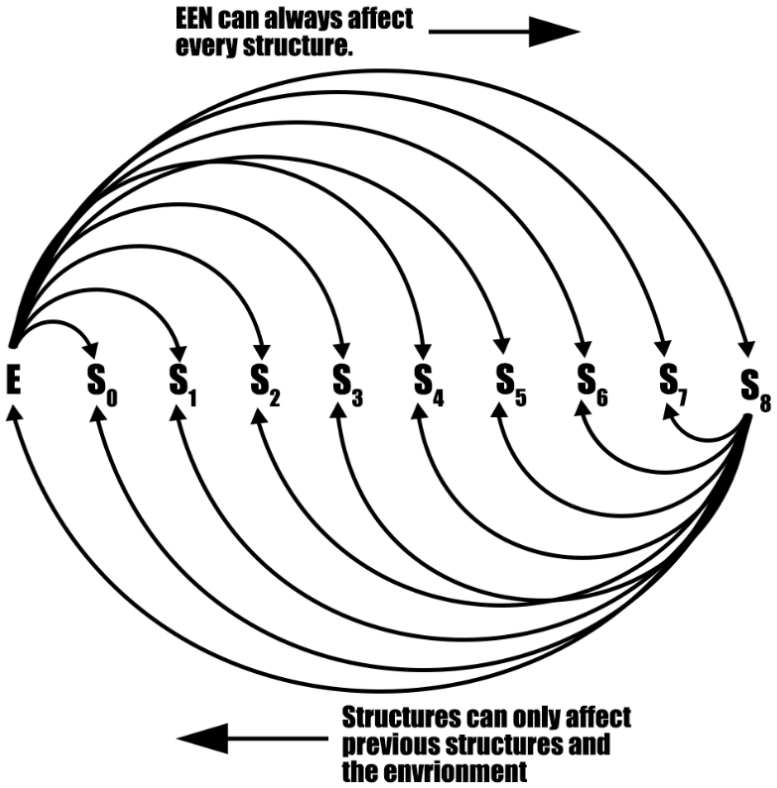
因果生态位

起初，我们只有一个环境能量稳态回路（EEL），其“目标”是将自由能降至最低。由于起初所有能量都是自由的，回路首先“想要”做的就是将能量困在结构中。它生成的第一个结构 S_0，必然具有因果力，使得环境能够重现它。要实际实现这一点，我们需要让 EEL 具备一项属性，用于详细说明每个时间步中它内部的每个结构应获得多少能量。默认能量值为 0。因此，要想存在，S_0 必须感知这一特定属性，并将其设为自身所需的自由能需求（FED）或更高。对于任意给定的结构，这里有三种情况：

1. 如果结构获得的能量低于其 FED，它就会不复存在。
2. 如果结构获得的能量恰好等于其 FED，它就能维持存在。
3. 如果结构获得的能量高于其 FED，且这部分能量没有被明确使用，那么除非有其他“事物”明确将其移走，否则这部分能量会继续被困住。
    a. 重要的是，该结构接收到的能量必须低于其自由能上限（FEL），这为结构所能包含的能量设定了上限。

回想一下，这种动态与我们前述关于能量稳态循环内部结构的推理是一致的。循环内部的结构是抽象化的能量稳态循环，它们始终保持在各自的稳态设定点，因此持续施加相同的因果力。我们假定，这些结构所属的能量稳态循环会将处于稳态之外的结构视为“失效的”，并因此将其解体。

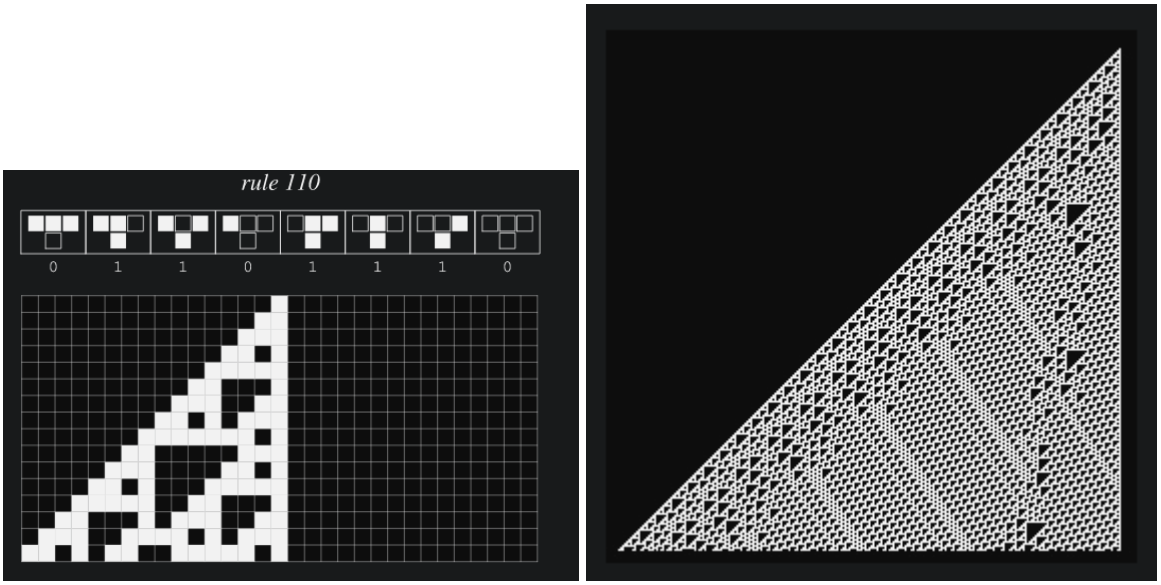
重要的是，S_0 **只能**与 EEL 交互，而不能与其他结构（S_n）交互。这是因为，不可能明确指定 S_0 与一个尚不存在的结构的未知属性之间的交互。然而，S_0 可以通过其对环境的影响，间接地与未来的结构交互。只能与此前生成的结构交互这一限制，对于之后生成的每个结构仍然成立，由此我们得到了一张展示每个已生成结构之因果生态位的图。因果生态位被定义为一个结构能够感知并影响的属性集合。



- 因果生态位图展示了结构只能影响环境以及此前已创建的结构。图中以结构 8 和环境为例绘出了箭头；环境始终能够影响每一个结构。

之所以如此定义结构，是因为替代方案无法处理，或者至少会更接近于归纳物理定律的实践。理由如下：为了生成一个能够明确影响未来结构的结构，需要有一种媒介（即通用的通信方法），能够支撑这些特定的规范。每个结构都“存在于”这种媒介之中，并通过它运行。为了更好地理解这一点，可以类比物理引擎。在物理引擎中，我们通常只关注牛顿物理学，换句话说，关注“物体 A 撞上物体 B 时会发生什么？”这一问题。由于底层物理规律是已知的，因此无需担心哪个物体先被创建，就能使这些物体正确地发生交互。描述力如何作用于物体并使其移动的物理定律是已知的。而在我们的情形中，物理定律并不为人所知，因此两个“物体”之间的交互必须通过代码明确定义，而这段代码定义了该物体的因果力。

明确指定交互对于可解释性至关重要。事实上，人们一直在尝试使用构成系统“物理定律”的基本通用规则，探索可能的开放式演化系统。其中三个值得注意的系统是n维离散元胞自动机、连续元胞自动机和超图。问题在于，我们不仅无法保证开放式进化，即使成功实现，也很难以人类能够理解的方式弄清楚正在发生什么。因此，尽管通过代码明确指定交互存在一定限制，但这对于理解系统正在发生的事情，以及从原则上逐步接近开放式进化，可能都至关重要。



- 1一种名为“规则 110”的一维元胞自动机展示了：通过在简单的初始条件上运行一条简单规则，可以实现多大的复杂性。尽管它起始于一个黑色细胞，并使用一条简单的规则来决定每个细胞如何改变颜色（它只依据相邻的细胞），最终却会产生不可预测且可能永无止境的复杂性。换言之，简单的规则可以产生任意复杂的行为。事实上，研究发现规则 110 能够实现“通用计算”，也就是说，与典型计算机一样，你可以使用规则 110 运行任何可能的计算机程序。这里的问题在于对程序进行编码和解释的困难。有趣的是，仅仅探索可能的简单程序空间，就揭示出了一些模式；在宏观层面上，这些模式与自然界中发现的模式相似。相关地，斯蒂芬·沃尔夫勒姆领导着一项研究计划，旨在通过探索可能的简单超图程序空间来发现物理学的基础理论；超图程序本质上是元胞自动机更为简单的版本。通过这一探索，沃尔夫勒姆研究公司已经能够从超图中以数学方式推导出狭义相对论的基本特征，例如洛伦兹变换，以及光速作为普适速度上限这一性质 [**56**]

- 左侧显示程序的规则和前16步，右侧显示前250步。我们从第一行的一个黑色细胞开始，将规则应用于该行中的每个细胞，以生成下一行。

- 来源：Weisstein, Eric W.《Rule 110》。摘自 MathWorld——Wolfram 网络资源。https://mathworld.wolfram.com/Rule110.html

结构规格

如我们的因果生态位图所示，我们总共生成了 9 个结构，但它们完全是抽象的。我们所知道的只是，每个结构至少都必须具备组装／拆解成本、因果力（及其相关成本）、有效半径，以及其在空间中占据的位置。我们还可以为这些结构赋予其他属性，这些属性可能适用于所有结构，也可能仅为某些结构所独有。

对于普遍属性，我们的设想是，应当能够以某种方式直接根据上述属性计算出这些属性，从而使它们始终与该结构的规格相关联。例如，结构的质量可以根据其能量成本，使用 $E= mc^2 \rightarrow E/c^2 = m$ 计算，其中 E = 能量，m = 质量，c = 光速常数，或宇宙速度上限。此外，位置还可以作为质量的函数进行调整，因为改变物体的位置需要消耗一定能量。可使用以下公式计算：W = (1/2) * m * v^2，其中 W 是对物体所做的功，m 是物体的质量，v 是其最终速度。将某个属性值改变“1”个单位所需的能量在此称为“扰动能”，对于从能量角度为任何属性奠定基础都至关重要，无论该属性是独有的还是普遍的。

独特属性的一个例子可能是“膜电压”。该属性至少应具有一个

默认值和一个扰动能量；不过，第三个可能同样重要的特征是其在结构因果力中的“位置”。这将允许通过扰动结构属性来改变结构的因果力。可以设想利用当前膜电压来决定是否发送消息。出于这一原因，我们将此类属性称为“因果属性”。类似于利用通过经验发现的方程来决定如何修改任意结构的位置，也可以实现用于描述修改膜电压难易程度的方程。因此，这类属性有助于生成一个能量封闭系统，使其不仅具有可解释性，而且还与我们的物理现实存在相似之处，从而有望成为一种强大的建模工具。不过必须注意，这并非没有代价：实现的复杂性将大幅增加。至于如何实际定义这些属性，我们希望将其留给自动化任务来完成。

自动结构生成

因果生态位图使我们能够以一种有原则的方式持续生成结构。我们可以在每个因果生态位中填充各种类型的结构。由于每个因果生态位只限制感知输入属性和受影响输出属性，因此，实现感知信息并影响输出属性的实际因果动力学由我们决定。为了减轻这项工作，我们借鉴了 ELM 使用大语言模型（LLM）的方式。实际上，我们创建了一个函数，该函数具有特定的输入（感知属性集合）、潜在的内部变量（因果属性）和输出（受影响属性集合）；我们希望利用大语言模型来补充输入和内部变量如何产生输出的具体细节。借助大语言模型，环境可以在每个生态位中创建一组具有各种因果动力学的结构——生态位内结构的数量上限由程序员决定。至于有效半径，可以使用大语言模型来确定，也可以由程序员选择根据某种特定算法或标准来分配有效半径。

除了论证环境的目标是最小化其自由能之外，本文还指出，实现这一目标的一种有效方式，是让环境同时测试不同的结构组合，看看哪些结构组合能够构成复杂的复合结构，从而最小化 VFE。那些无法充分最小化 VFE 的组合会在竞争中落败，因此未来被创建出来的可能性也会降低。

在讨论结构如何以这种方式相互混合和匹配之前，重要的是先讨论一种自动为此类结构分配能量成本的机制。

自动分配能量成本

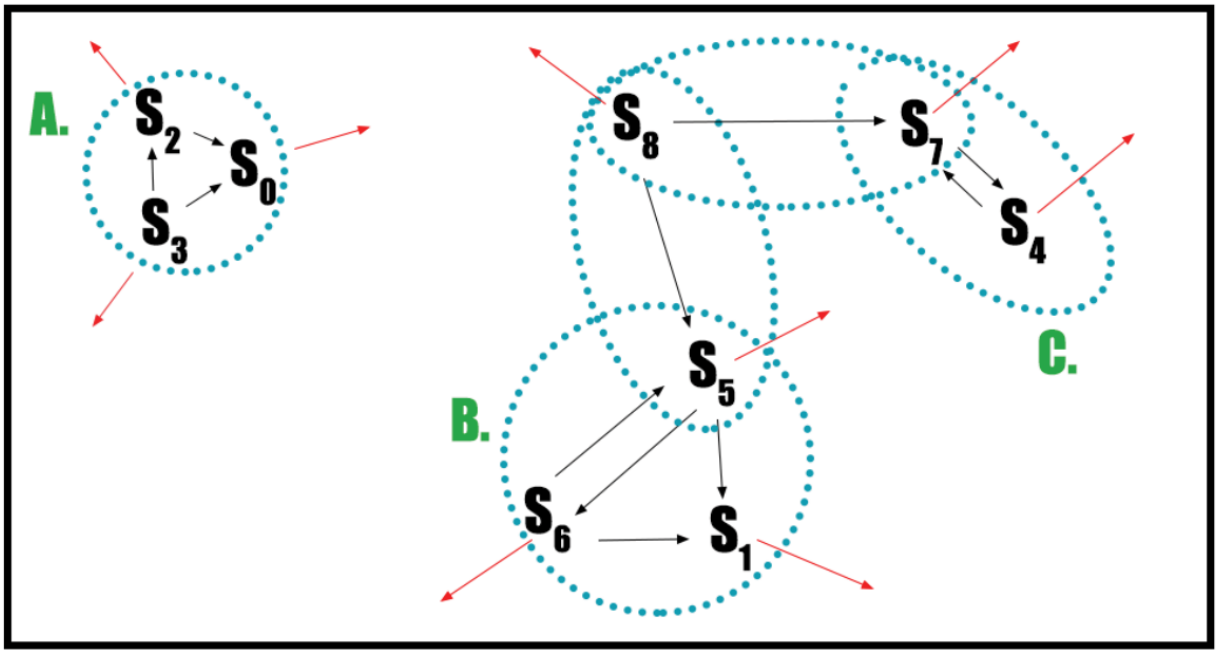
此前，我们通过直觉或建模实验所得的经验数据，为结构分配能量成本。需要注意的是，必须谨慎行事，不能违反能量既不能凭空产生也不能凭空消失这一约束。考虑到这一点，我们可以采用与之前类似的方法，但也可以选择另一种能够实现自动化的方法。自动分配能量成本的好处不仅在于省去这项工作，还能确保各种创建出来的结构在能量方面保持一致。换言之，因果力相似的结构具有相似的能量成本。这为我们提供了一种“底层物理规律”，我们的能静态程序必须遵循这一规律。值得注意的是，这一技术同样适用于能量开放的能静态网络系统，但在建模场景中应用起来可能不那么直接。

为了自动为结构分配能量成本，我们首先提出这样一个观点：结构组装与拆解的成本大小相等、方向相反。此外，结构组装的成本等于因果力的成本。由于因果力只是一个函数，其函数体是由大语言模型生成的代码，因此我们可以获取该代码，并将其转换为“抽象语法树”（AST）。AST 是一段代码的简化表示，通常由编译器或解释器构建，用于高效地分析和处理源代码。撇开技术细节不谈，AST 使我们能够计算定义因果动力学的代码中包含的基本操作或原语的数量。掌握所选编程语言中的所有可能操作后，我们便可以为每种操作分配能量成本，进而计算因果动力学所需的总能量。在不同编程语言内部以及不同编程语言之间切换这些成本，构成了一个颇为有趣的研究领域。至于因果属性，则必须使用其扰动能量函数来计算其成本。此类函数接收属性的当前值，并返回修改该值所需的能量

。随后，可以从 0 开始，在属性逐步增大至默认值的过程中迭加各项成本，从而计算该属性的成本。因此，因果属性代表了存储的能量。

结构重组 - 信息和智能都是物理的。

正如“能量稳态循环、网络与巢穴”一节中所述，能量稳态循环这一概念旨在表达这样一种观点：可靠地产生某种结构所需的信息，是由这样的循环以物理形式承载的。此外，我们提出，结构本身确实就是一个能量稳态循环；其可能变化的架构之所以被抽象掉，是因为只有当该结构处于某一种特定的静止状态时，才会被视为“稳定”，或对其嵌入其中的循环有用。本文认为，能量稳态循环的动力学应被视为物理定律所决定的一种普遍结果；这些定律规定了各个尺度上组成结构之间哪些关系是稳定的。因此，能量稳态循环应被理解为结构间的关系，其稳定性取决于：（1）结构之间的相互作用；（2）循环所处的特定局部环境。这里的“环境”既指 EEL，也指在这样的意义上“充分处于”相关能量稳态循环之外的所有其他结构：能量稳态循环内部的结构与循环外部的结构之间，仅存在稀疏、微弱或其他不稳定的关系。这些关系的形成方式很简单。当一个结构处于另一个结构的有效半径之内时，二者便形成一种关系。这种关系构成了对参与其中的结构的信息的具身承载。下图展示了这类稀疏关系可能呈现的样貌。



- “环境能量稳态回路（EEL）内部已创建并彼此建立关系的结构”示意图。红色箭头表示结构因果力对环境的影响，黑色箭头表示结构因果力对其他结构的影响。蓝色虚线表示能量稳态循环，应理解为结构之间的关系。只有当结构彼此产生影响时，结构之间才存在关系。

- 回路 A 中的结构与任何其他结构都没有关系。回路 B 和回路 C 与结构 S_8之间的关系较为稀疏，因为它们受到 S_8 的影响，却不会影响 S_8。

如前所述，每个能量稳态循环都有一个“能稳态窗口”，由“稳态窗口”、“行动窗口”和“行动变化概率窗口”组成。不过，我们不再将循环限制为只能创建它们曾接触过的结构，而是允许循环创建 EEL 已经“发明”过的任何结构。这进一步模糊了循环与环境之间的界限，因为每个循环本身都被视为一种微型 EEL。因此，使每个能量稳态循环与其他循环充分

不同的是，它们对于给定能量水平时各自将采取何种行动，具有不同的概率分布。当结构彼此远离到一定程度、所采取的行动拆解了某个结构，或者流经循环的能量不足以维持其动力学，或超出该循环的耗散能力时，关系便会瓦解。

定义能稳态窗口同样是我们希望自动完成的事情。为此，我们再次审视抽象语法树（AST）。抽象语法树使我们能够计算任意结构的自由能需求（FED），同时也使我们能够计算任意结构的自由能上限（FEL）。这两个值分别定义了结构维持自身存在所需的能量，以及结构能够耗散的最大能量。了解能稳态窗口中每个结构的这些值及其连接关系后，我们就能创建一个函数，用于计算能稳态窗口的能量下限和上限。通过这种计算，我们的能稳态下界等于维持系统所需的最小能量，而能稳态上界则等于通过网络能够耗散的最大能量。超出这些界限后，结构会丧失，循环也随之瓦解。

总结
我们从一个环境能量稳态回路（EEL）着手。它由一个“自由能”池、一个用于在“因果生态位”中生成因果结构的大语言模型（LLM）、一种用于为每个结构分配能量成本（FED）和限制（FEL）的能量映射，以及一段关于结构如何像能量稳态循环那样建立、破坏和调整关系的描述组成。至于具体实现方式，这些元素都具有灵活性。EEL 的初始自由能量、因果结构的生成方式、能量成本映射到逻辑代码的方式，当然还有结构重新组合并调整其关系的方式，都可以有所不同。具体实现并不代表 Ensoul框架，只反映了执行该框架的特定方式。话虽如此，这里的核心理念是开发能量封闭系统，使其生成的结构变得愈发有趣、智能，并且按照“开放式演化与组装理论”一节中的一般定义，愈发“有生命力”。

量子计算表明，在物理上创建完全能量封闭的系统可能并不现实，甚至并非理想之选。然而，这类系统的近似形式或许能够成为强大的途径：既更轻松地在物理世界中生成智能技术，也能作为模型，帮助我们更深入地理解热力学开放系统的演化与动力学。在着手创建此类设备的物理形态之前，首先在计算机中展示它们的能力，并利用这类模型系统为设备的创建提供指导，是十分重要的。

我们在本文中实际上所做的，是通过在代码之上构建一种底层的“能量物理学”，使代码获得物理形态。换言之，“结构”就是被具体呈现在能量受约束的虚拟领域中的代码。这与这样一种观点相一致：物理世界中的一切都可以被视为在“进行计算”，因此都可以通过通用图灵机所能执行的计算来描述。这种观点被称为泛计算主义 [4]。虽然我们无法确定物理世界中的一切是否都可计算，但同样无法确定，除了使用数学和计算工具来理解它之外，我们是否还有其他选择。

到目前为止，讨论的重点一直是在计算机内部采取哪些措施来创建 SOULS。然而，这些模拟本身对于软件技术的发展具有重要意义；不过，使用传统冯·诺依曼架构 IOS 对能量封闭系统进行模拟，会严重低估所用材料的真正计算能力。这种未被充分利用的计算能力，正是我们在《IOS Illusions: Motivating Modeling With Enerstatic Loops》中所称的“高度可变行为”或“噪声”。工程师一直煞费苦心地试图避免材料具备这种特性，但只要采用恰当的方法，这种特性实际上可以作为有用的计算过程发挥作用。将噪声视为计算资源的系统被称为“热力学计算机”（TC）。这些计算机在物理现实中就是 SOULS，真正让我们更接近 SOULS 定义中的“超低功耗”部分。

热力学计算与能静态网络

热力学计算机被设想为一种“经过工程设计的、多尺度的复杂系统；当暴露于外部势（输入）时，会在短期内自发地在不同状态之间转换

同时在长期内不断优化其组织形式，二者都是热力学驱动的内在适应过程的一部分”[33]。这类似于我们所描述的能静态网络的行为，只不过是以物理形式呈现。本文认为，收集和研究这一一般概念的各种表现形式，将有助于理解嵌入式能动性这一问题。在本节中，我们将简要讨论热力学计算机、热力学计算与能静态网络之间的关系，最后讨论嵌入式能动性。

热力学是物理学的一个分支，研究热和功等各种不同形式的能量之间的关系。它最初是为了优化从蒸汽机中提取的能量而发展起来的[9]。热力学的这一应用聚焦于“平衡态”，在这种状态下，系统的状态均匀且在宏观上保持不变，例如一池静止的水。事实上，热力学得出的结论之一是，整个宇宙都在朝着这种平衡态演变，这就是所谓的“宇宙热寂”。尽管有这一结论，许多物理学家开始相信，热力学也是推动我们所知的生命形成的主要力量[47, 46]。这推动了“非平衡热力学”的发展，人们认为生命以及许多其他远离平衡态的复杂系统似乎都运用了这种热力学。尤其是，关于涨落耗散定理和信息热力学的研究持续带来新的见解，帮助人们理解生命以及热力学计算机的发展。

热力学计算面临的问题之一，是缺乏能够恰当表述热力学计算的统一通用模型系统。尽管已有一些模型系统，例如霍普菲尔德网络、玻尔兹曼机器、热力学神经计算模型和热力学神经网络，但这些模型被视为仅仅是“一小组有趣且具有启发性的装置与架构[9]”。虽然这一研究领域仍有大量工作有待开展，但本文认为，能静态网络或许足够通用，可以作为这样一种统一的通用模型系统。这类系统对于评估核心理论概念（如涨落耗散定理和信息热力学）、展示最适合采用此类方法解决的问题，以及为热力学计算机的开发提供有益参考，都至关重要。

推测能静态网络可能充当此类模型的依据，在于其组成部分——能量稳态循环——本质上是对热力学开放系统的极度简化抽象；通过采用实测值并整合根据实证推导出的数学方程，可以直接扩展其细节层次。

鉴于 SOULS 的存在，无论它们是虚拟的、物理的、工程化的，还是进化而来的，以及控制或利用上述 SOULS 的需求，人们需要一个适当的框架，或一种思维方式，来实现这一点。这让我们回到了 TAME——普遍心智的技术路径——它提供了许多这样的思想和信念，以实现这一目标。

驯服 SOULS

虽然任何“事物”的目标都是存在这一点显而易见，但很明显，“事物”似乎也拥有“更高层次”的目标，而这些目标看起来与存在本身几乎没有什么关系。例如，一个选择登上珠穆朗玛峰顶峰的人，或选择在灼热阳光下穿越死亡谷参加一场135英里赛跑的人，似乎并没有在采取通向持续存在的最优路径。

具有欺骗性的高层级目标

本文认为，对存在的追求可能具有欺骗性，这或许类似于我们在讨论双足机器人的训练时所说的通往特定目标的路径具有欺骗性。也就是说，通往存在的路径似乎是发散的，而非收敛的。不变多尺度 TAME 方法为描述这种表面上的不一致提供了一种有趣的方式。

首先，任何“事物”都由一些以存在为目标的部分组成。对于许多事物，尤其是我们认为“更有生命力”的事物而言，这些部分之间并非紧密相连。换句话说，每个部分都没有与其他所有部分直接形成因果连接。考虑到“自我”的连贯性取决于各部分共享应力的程度，某些部分可以被划分为属于不同的“事物”，或者换句话说，在某种程度上构成独立的“自我”。一个容易理解的例子是生物学的

有机体由细胞（部分）组成，这些细胞聚集在一起，形成不同的组织（事物）；组织又以不同方式聚集，形成器官，依此类推。如前所述，当资源有限时，实现存在目标必然会导致其他“事物”面临资源限制，因为可供利用的自由能更少了。因此，彼此不直接共享应力意味着，任何“事物”的各个部分如今都有可能为了争夺资源而相互竞争，尽管它们最终属于同一个“事物”。

然而，存在竞争的可能性并不意味着竞争必然正在发生。换言之，尽管细胞集体具备严格竞争并采取会损害整体集体的行动的能力，持续存在的细胞集体仍通过合作与竞争的结合达成共识，我们在此将其称为“竞合”。竞合动力学所产生的共识，在金融市场和企业中得到了相当贴切的体现：个体以及由个体组成的群体仅根据自身的局部利益采取行动，却能够实现稳定的市场价格，并确保所有参与者的收支保持平衡。竞合存在于集体智能的各个尺度，从细胞器、细胞、组织到器官，等等。根据 Stephen Grossberg 博士关于此类竞合系统的研究，各部分在特定决策上的自由程度，与能够作出的全局决策的规模之间存在权衡。这一限制可能导致网络中存在多个选择，其中某个新颖的选择可能受到偏好；也可能导致不同决策或不同部分之间出现振荡，或发生“跳跃”，使当前“赢得”竞争的决策或部分发生变化 [**9, 18, 17, 19**]。这两种现象都会促成有趣而复杂的适应性行为，因为相关集体会采取行动，而在有利情况下，这些行动能够使集体得以生存。Grossberg 博士证明了一个关于这种动力学的定理，从而可以设计出此类系统，该定理被称为“适应水平系统定理”。作为一个极其简单的例子，可以这样理解：对于我们的登山者而言，到达珠穆朗玛峰顶峰这一行为可能会损害集体中某些细胞群体，却有利于其他细胞群体的生存。从这一行为中获益的细胞群体正在“赢得”竞争；但当登山者下山后，另一些细胞群体又会“赢得”竞争，由此实现各方参与者之间的收支平衡。

现在，我们已经了解到，哪些因素可能在许多情况下导致“事物”表现出看似与存在无关的行为。本质上，这取决于“事物”内部各部分之间的竞合关系；根据其构成部分共享应力的程度，这些部分可以被称为“部分”。接下来要做的，就是弄清楚如何识别集体所作决策的目标或目的。这种努力使我们能够更好地描述、设计并与这类集体协作。

驯服 SOULS 与认知能力测试（CCTs）

知觉控制理论（PCT）是一种控制理论，专门用于理解自然产生的控制系统，即 SOULS。知觉控制理论（PCT）认为：
“自然控制系统只能围绕其行动（或独立事件）对其输入（广义定义）的影响来组织，因为其输入包含了其行动所产生的、任何可能对控制系统产生影响的后果 [**43**]”。

Powers 用寥寥数语精辟地描述了这一概念：“行为就是对感知（输入）的控制”。换句话说，采取行动是为了确保这个“事物”感知到它“想要”感知的内容。在知觉控制理论（PCT）中，这种期望的感知称为“参考信号”。至关重要的是，人们认为它是在内部生成的，而不是像大多数人工控制系统那样由外部生成。以这种外部性为例，恒温器试图调节的温度值取决于用户提供给它的参考信号，而且通常不能由恒温器自身修改。对于自然控制系统，参考信号始终可以由“事物”自身修改。再来看看外场手问题的一种非分析性解决方案——“光学加速度抵消”。在这个例子中，自然控制系统，也就是外场手，想必会自行选择要控制的变量。同一个控制系统还可以致力于控制许多其他变量，以完成不同的任务，因此它在认知连续谱上的位置会比简单的恒温器靠右得多。

受控变量测试（TCV）

推断任何控制系统的受控变量，所采用的程序相同，无论该系统是自然的还是人工的。在知觉控制理论（PCT）中，这一程序称为“受控变量测试（TCV）”。TCV 可以用来识别所观察的系统类型（IOS 还是 SOULS），并在存在受控变量时，识别 SOULS 正在控制的变量。TCV 是一个反复进行的经验性过程，通过这一过程，人们检验关于某个变量是否受到控制的假设。假设的形式如下：“如果变量 X 是受控变量，那么变量 X 受到扰动时所受的影响应小于通常情况。”恒温器控制温度就是一个直观的例子。假设空调开着，在寒冷的日子里打开窗户会使温度缓慢下降，而恒温器很快就会启动加热机构来抵消这一变化。如果恒温器并未将温度作为受控变量，这种情况便不会发生；而这正说明温度受到扰动时所受的影响“低于通常情况”。TCV 的一个重要应用，是确定“光学速度控制（COV）”比“光学加速度抵消（OAC）”更能解释生物体如何解决外场手问题 [**33**]。速度和加速度这两个变量密切相关，因此很难在扰动其中一个的同时不影响另一个。为使判断更倾向于某一方向，研究人员建立了两个模型控制系统，分别只控制这两个变量中的一个。OAC 模型解释了生物体拦截物体路径中 75% 的方差，而 COV 模型解释了 93% 的方差。这表明，在拦截物体时，实际受到控制的是光学速度，而非光学加速度。TCV 通过确认没有任何变量受到该系统行为的控制，排除了我们所研究的系统属于 IOS 的可能性。对于像“古怪恒温器”这样的系统——它只有在 50% 的时间里才通过输出调节温度——结合使用统计方法与 TCV，可以帮助我们判断该系统在 IOS 与 SOULS 连续谱上的位置。

虽然受控变量测试（TCV）能够帮助我们确定系统在执行某项特定任务期间所控制的变量，但它无法告诉我们控制系统具备控制哪些变量的完整能力，也无法告诉我们应如何修改系统受控变量的参考信号值，或如何诱导系统在不同受控变量之间进行切换。这类探究或许会促使我们分别提出“变量可控性测试（TVC）”和“控制切换测试（TCS）”。下面我们将简要讨论它们可能如何运作。

变量可控性测试（TVC）

变量可控性测试（TVC）试图回答这样一个问题：“对于我能够测量的变量，系统能够感知其中哪些变量？”这种表述清楚表明，观察者主要负责正确推断系统的认知能力。与受控变量测试（TCV）一样，变量可控性测试（TVC）从关于某组可测量变量的假设出发，并为每个变量赋予一定的可信度。测试包括传感器识别，以及随后推断控制系统可能进行哪些感觉划分。换言之，就是识别传感器的位置和类型，并确定这些传感器作为一个整体可能感知到什么（例如光检测（小尺度），与对整个物体的感知（大尺度）之间的区别）。一个简单的例子是：根据某种生物拥有眼睛这一事实，推断它可能能够检测单个物体。尽管这并非易事，但这项任务对于绘制系统“认知光锥”的空间维度十分重要；认知光锥是系统为实现特定目标而具备的时空计算能力。时间维度与空间维度密切相关。在变量可控性测试（TVC）中，我们最终关注的是系统将能量用于控制何种类型的可测量变量，以及扰动在时空中距离多远时仍能引发系统行动的变化。简而言之，变量可控性测试（TVC）应从理论最大值出发，通过开展实验测试来逐步缩小可能性范围，从而着力绘制系统的认知光锥。

控制开关测试（TCS）

控制开关测试（TCS）旨在确定应采用何种方法来修改系统正在控制的变量的值和/或身份。TCS 与 TA ME 提出的“可说服性轴”概念有关。请回想一下，系统在可说服性轴上的位置代表了要让系统按照期望运行，所需进行干预的复杂性层级。系统在可说服性轴上越靠右，需要耗费的能量就越少。换言之，一个良好的控制开关最终应使人能够尽可能少地耗费能量。如

此，TCS 应测试从最高潜在认知能力开始、处于不同认知能力层级的开关；其中最高潜在认知能力是通过 TVC 推断出的认知光锥。简而言之，TCS 会问：

1. 我能否通过某种媒介，使用某种语言与该系统进行交流？ 2. 我能否识别出可用于训练该系统的奖励？ 3. 是否存在可以通过某种物理手段修改的设定点？ 4. 我真的必须把它分解成各个部分，再从头开始重新组装，才能让它完成我想要的事情吗？

这里的重点是避免低估一个系统的智能。一个极佳的例子展示了控制开关的力量：通过修改生物电所体现的设定点，可以实现形态控制 [31]。Michael Levin 博士的实验室及其合作者通过几个关键实例，确立了细胞的生物电状态能够指导其集体目标形态。首先，需要了解的是，细胞内部通常比细胞外部带有更多负电荷。细胞内部的电荷或电压是可以测量和可视化的。这使我们能够观察某些药物或其他干预措施对部分细胞集体生物电状态的影响。我们将提到的 Levin 的第一个实验是在涡虫身上进行的。涡虫是具有双侧对称性和神经系统的复杂生物体。它们还拥有惊人的再生能力。涡虫可以被切成两半，随后每一半都会发育成完整良好的涡虫。研究已经证明，小至动物体积 1/279 的一块涡虫组织，也能重新生长为完全具备生存能力的成年涡虫 [38]。通过观察细胞的生物电状态，人们实际上可以操纵“本应成为”尾部的细胞，使一只被切断的涡虫长出两个头，而不是一个头。人们已经以这种方式培育出双尾涡虫、长有其他涡虫物种头部的涡虫，甚至培育出长有头部的涡虫，而这种头部在迄今为止我们所知的涡虫自然界中可能根本不存在。重要的是，一旦完成这些生物电操纵，涡虫在再次被切断时会记住这种状态。换句话说，一只双头涡虫被切断后，会生成两只双头涡虫。尽管基因组完全没有被修改，情况依然如此。这代表了针对生物学的有机体的一种强大控制开关。同样的情况也出现在蝌蚪身上：通过生物电操纵，可以让蝌蚪的尾部长出一只眼睛。蝌蚪尾部出现眼睛并不需要进行基因修改；蝌蚪的细胞实际上会根据一个相对简单的局部信号，自行组织起来，使这只眼睛具备功能 [41]。研究人员还使蝌蚪原本的眼睛失明，以证明它仍然可以依靠尾部的眼睛导航 [6]。关于细胞生物电状态作为控制开关的进一步证据，来自所谓的 Xenobot 的创造。Xenobot 是 Levin 与 Josh Bongard 博士合作创造的。Xenobot 本质上是这样形成的：青蛙的皮肤细胞在发育过程中脱离了通常来自身体其他部分的电信号。脱离这些信号后，这些细胞会凝聚成一种原始有机体；这种原始有机体通过物理移动周围的其他细胞，使其形成类似的原始有机体，从而进行自我繁殖。这些现象清楚表明，细胞的生物电状态充当了控制其行为的设定点。事实上，Levin 证明，通过迫使癌细胞在电学上重新融入集体，可以有效治愈某些癌症 [8]。

识别系统的认知光锥和控制开关都是艰巨的任务，但其回报显然值得付出相应的努力。只有在绝对必要时，才应通过重新布线系统的硬件来控制系统。自然控制系统所表现出的能力应当清楚地表明，为了确定以最高效的方式理解并利用其行为以实现特定目标，必须适当评估构成自然控制系统各部分的智能。重要的是，这些测试都可用于理解和开发热力学计算机。使所有这些测试更加严谨并能够形式化，是 TAME 的目标，也是本文的重要动机。一个能够捕捉热力学驱动的复杂适应系统之自组织动力学的模型，将成为理想的试验平台，用于理论计算、经验测试/统计学的开发，以及更具生物学启发——或者更准确地说，更具热力学启发——的技术进步的开发。现在，我们回顾驱动所有这些行为的期望感知，并讨论未来的发展方向。

讨论：研究方向、推测与哲学

总结 Ensoul

“鉴于我们看到神经元在有机体的整个生命周期中持续存在，那么为了做到这一点，它必须非常擅长做什么？”——本文认为，这件事就是“管理其能量水平”；在这种情况下，

假设存在任何其他目的，对其行为的分析都会受到阻碍。然而，信息处理确实似乎正是神经元所做的事情。那么，是否有可能构建这样的系统：其中的信息处理，或“工作”，仅仅作为某个“事物”管理其能量水平的结果而完成？这一概念是否也能更普遍地关联到我们在物理现实中观察到的现象？——本文对这两个问题都给出了肯定的回答。这一概念的推广促成了受热力学启发的抽象概念——能量稳态循环。其核心驱动力在于：任何存在并持续保持其存在的“事物”，都必然参与某种能够维持这种持续存在的过程。换句话说，它必须具有因果力，而这种因果力要么嵌入某个其他“事物”的自我证据化行为之中，要么自身表现出某种程度的自我证据化行为。这一框架意味着，任何持续存在的事物，都可以很好地理解为一个控制系统，其中负反馈维持着那些本来会瓦解的状态，而这种动态机制是由构成该系统的各个部分内在地产生的。

遵循这一思路，本文倾向于泛计算主义，即认为这类因果力总能以形式化方式表述，并且可以计算。然而，由于需要同时考量能量与信息处理，手工编写此类代码并非易事。此外，开放式进化的研究发现清楚表明，要实现预期结果，优化（收敛）技术并非总是最佳途径。值得庆幸的是，enerstatic loops 凭借其存在的内在驱动力和信息处理能力，非常适合开放式进化。将能量成本映射为代码（这一过程可以称为“去抽象化”或“物理化”），运用大语言模型，以组装理论作为衡量进化系统复杂性的基础，以及自由能原理所提出的信息对称性，共同表明，探索“能量封闭系统”是理解并生成基于能量稳态循环的热力学开放系统模型的一条前景广阔的途径。鉴于这类能量封闭系统及其生成的能量开放系统具有抽象且独立于基底的特性，“自组织智能超低功耗系统（SOULS）”这一术语似乎很适合用来描述这两类实体。

对 SOULS 等“事物”的理解，表明我们有必要以一种基于存在论的多尺度视角来看待现实；这种视角聚焦于不变的概念，并将其一致地应用于每个尺度上的每一种“事物”。普遍心智的技术路径（TAME）是采取这种视角的重要框架。因此，“智能”“认知”“目标”“应力”“自我”“生命”和“反馈”等概念都被定义为一个连续谱，而不是离散的类别。我们关注的不是这种“事物”是否具有这些属性，而是它属于“何种类型”，以及“程度如何” [30]。这对于理解复杂系统至关重要，尤其是有助于我们了解这些概念如何逐级扩展，以及这种扩展在宏观层面会呈现出什么结果。

显而易见，技术进步对人类的影响正以指数级速度不断加剧，目前看来仍没有放缓的迹象。这些进步带来的结果往往难以预测，既有正面影响，也有负面后果。因此，本文所述技术的开发旨在实现两个目的：（1）创造可供实际应用的实用技术；（2）将其作为试验平台，以便更好地理解并减轻负面后果。理解此类技术的后果并非易事，但我们不仅要能够构建它，也必须能够控制它，这一点至关重要。由于不了解当前技术所带来的后果，我们已经遭遇了隐私丧失、社交媒体成瘾以及前所未有的政治影响力等问题。与此同时，有充分理由认为，技术进步和全球互联互通的增强显著提升了整体生活质量，包括贫困状况下人们的生活质量。因此，在开发 SOULS 的同时构建 TAME 这样的框架，既是关乎人类生存的迫切要求，也是通往理想未来的必要之举。我们越快加深理解，就越有可能主动塑造技术进步所带来的后果。由此，我们开始思考重要的未来研究方向。

未来方向

简而言之，未来方向包括：

1. 使用传统的冯·诺伊曼计算机来实现能静态程序，例如通过现场可编程门阵列或其他灵活的经典计算方法。2. 开发能静态程序的专用版本，作为试验平台，通过推进理论／数学框架来助力热力学计算机的开发。

3. 开发能静态程序的专用版本，以助于理解生物学系统；本文所述的能量静态神经网络就是一个例子。4. 为能量封闭的能静态程序开发严格的智能、有趣性和“生命性”（复杂性）度量，使不同程序能够进行比较和自主分析。5. 开发 TAME，更具体地说，开发严格的认知能力测试（CCTs），（即智能／认知的存在性度量、受控变量测试（TCV）、变量可控性测试（TVC）和控制开关测试（TCS）。6. 最后，如有必要，创建替代方法，不再使用大语言模型在能量封闭系统中生成结构。

推测与哲学

本文的内容旨在汇集可能有助于推进科学与技术理解和发展的思想。为此，本文采纳了实体一元论、过程哲学、泛计算主义，以及扩展版的强生命—心智连续性论题。哲学至关重要，因为我们看待世界的方式会影响我们在世界中的行为，其中也包括我们开展科学研究的方式。因此，我们现在将逐步探讨推动 Ensoul 发展的部分哲学思想。

实体一元论认为，一切事物都由“同一种东西”构成，这似乎引出了这样的问题：“如果一切都由同一种东西构成，为什么我们会看到彼此不同的‘事物’？”以及“这些事物是如何产生的？”。本文除了回答“事实就是如此”之外，并未提供其他答案，但提出了物理上和逻辑上有依据的指导原则，说明如何模拟这类“事物”的行为，以及如何利用这类“事物”演化系统。以下论述提出了一种继续探讨的思路：

1. 物理学告诉我们，$E=mc^2$，，换言之，每一个“事物”都是由能量构成的。2. 我们通过因果关系来理解“事物”的行为。3. 如果一个“事物”没有任何影响，就不能说它存在。换言之，存在就是成为某种影响的原因（埃利亚原则）。4. 同样，如果一个“事物”的存在没有任何原因，就不能说它一直存在。换言之，持续存在就是成为某个原因的结果（持续性原则）。5. 因此，我们应将“持续存在的事物”建模为具有持续的原因，而这些事物本身也可能是这些原因的起因。换言之，就是对它们的存在性因果力和自我证据化因果力进行建模。

这引出了这样一种连续图景：各种“事物”具有不同程度的所谓“存在性反馈”。当一个“事物”造成的影响也是其自身原因的一部分时，就会出现这种情况。事实上，如果并非如此，那么必定是由其他事物造成了它，而那个事物本身也必定由自身或其他事物造成。因此，我们提出这样的假设：能够造成自身存在的事物，其存在更为普遍／持久。这些“事物”可以很好地理解为负反馈系统或控制系统。这些自然控制系统的有趣之处在于，它们调节的是自身特有的动力学或因果力，而非某个特定的可测量变量。这种观点与过程哲学十分契合；过程哲学认为，没有任何“事物”是静止的，一切都在不断变化。本文将这种必要的“变化”动力学与埃利亚原则及因果律相结合，以论证存在性循环因果关系或存在性反馈回路是对这类系统进行抽象的恰当方式。为此，本文借鉴了泛计算主义和连续性论题本身。

泛计算主义与过程理论一样，认为物理世界可以被理解为一个过程；不过，它具体主张这一过程是计算性的。也就是说，这是一个可以在计算机上模拟的因果过程。本文之所以倾向于这一观点，有两个原因：（1）如果我们要理解现实的本质，就必须使用因果语言，因为因果关系在计算设备中的自动化，已经以我们从未想象过的方式扩展了我们的认知；（2）这种方法或许能够通过证明那些通常被描述为“涌现”的属性，确实无法被描述为由构成该现象的基本组成部分以计算方式实现，从而提供有力证据，说明泛计算主义不可能是充分的。

由于倾向于认为，这些属性确实可以通过构成宇宙的基本“事物”的动力学以计算方式实现，本文接受了强生命—心智连续性论题 [**25**]。强生命—心智连续性论题认为，生命的起源与心智的起源是同义的。换句话说，一旦有了生命，就有了心智。本文试图将智能和认知等熟悉属性操作化为依赖于某个“事物”的自我证据化因果力。本文所倡导的努力正是建立在这些基础之上。

当前面临的挑战，是将存在性反馈回路这一抽象概念用作工具，以理解这些在操作层面定义的属性如何逐步扩展为我们所熟悉的各种形态，同时探索如何利用此类回路的动力学来创造新的技术，并增进对可能存在的生命形态的理解。为此，将演化方法与创建虚拟和物理空间中的能量稳态循环、网络和巢穴的传统方法相结合，或许会成为一个富有成果的探索领域。

最后还要提醒一点：切勿将地图误认为疆域。哲学、关于现实的理论和模型，尽管具有预测能力或令人信服，却并不等同于现实。我们所处的境况似乎是，我们无法真正知晓感官背后的原因究竟是什么[**21**]。哲学、理论和模型归根结底只是我们为那些可能存在的隐秘原因编织的故事。SOULS 这个术语碰巧很适合指称任何遵循过程哲学和实体一元论所阐述的故事、并持续维持其存在的“事物”。因此，我们可以追问：从形而上学角度看，SOULS 这一概念能否为日常用语中的“灵魂”赋予更严谨的定义？从哲学角度来说，Ensoul 似乎意味着灵魂是任何随时间持续存在的因果力。因此，像 Ensoul 这样运用存在性反馈回路的方法，可以被视为一种可实现的哲学模型，借此或许能更好地理解这一哲学叙事。故事体现了我们对现实中正在发生之事的主观信念，因此，我们讲述的故事会引导我们在生活和科学中的行动。所以，尽管承认这些故事并不等同于现实十分重要，但如果我们相信这些故事，它们同样至关重要，因为这些故事会深刻影响我们在世界中采取何种行动。有趣的是，“心理学”一词最初被定义为“灵魂研究”，后来才更普遍地被理解为“心智研究”。事实上，“psychology”一词由希腊语“psych⁻e”（意为“呼吸、生命原理、生命、灵魂”）和“logia”（意为“言语、词语、理性”）组合而成。Ensoul 的目标之一，是认真看待这样一种观点：心理学所研究的一些属性，例如智能和认知，并非只存在于人类心智中，而是存在于所有“事物”或 SOULS 之中。因此，心理学中的方法和思想，例如 PCT 和强化学习，有助于理解任何“事物”的本质，而不论该“事物”由什么构成。这或许意味着可以扩展或重新诠释心理学，称之为“新心理学”，也就是“对 SOULS 的新研究”。我们主要关注的是任何“事物”可测量的基质不变且操作性定义的属性（即智能、认知、压力水平等），以及这些属性如何逐步扩大，或者在何种情况下（如果存在这样的情况）可以说某些属性会“涌现”。由于本文主要采用建模或构建的方法来理解 SOULS，因此该框架被命名为“Ensoul”，这是一个动词，意为“赋予灵魂”。

参考文献

[1] 操纵自由意志信念及其后续影响：一项元分析。2022年。[2] Stephen Adams、Tyler Cody 和 Peter A. Beling。逆向强化学习综述。人工智能评论，55(6):4307–4346，2022年2月。[3] Subutai Ahmad 和 Jeff Hawkins。稀疏分布式表征的属性及其在层级时序记忆中的应用。2015年。[4] Neal G. Anderson 和 Gualtiero Piccinini。泛计算主义与物理系统的计算描述。2017年2月。[5] Saurabh Arora 和 Prashant Doshi。逆向强化学习综述：挑战、方法与进展。人工智能，297:103500，2021年8月。[6] Douglas J. Blackiston 和 Michael Levin。头部以外的异位眼为爪蟾蝌蚪提供用于光介导学习的感觉数据。实验生物学杂志，216(6):1031–1040，2013年3月。[7] Seville Chapman。接棒球。美国物理学杂志，36(10):868–870，1968年10月。[8] Brook T. Chernet、Chris Fields 和 Michael Levin。长程缝隙连接信号传导控制非洲爪蟾胚胎中由癌基因介导的肿瘤发生。生理学前沿，5，2015年1月。[9] Tom Conte、Erik DeBenedictis、Natesh Ganesh、Todd Hylton、John Paul Strachan、R. Stanley Williams、Alexander Alemi、Lee Altenberg、Gavin Crooks、James Crutchfield、Lidia del Rio、Josh Deutsch、Michael DeWeese、Khari Douglas、Massimiliano Esposito、Michael Frank、Robert Fry、Peter Harsha、Mark Hill、Christopher Kello、Jeff Krichmar、Suhas Kumar、Shih-Chii Liu、Seth Lloyd、Matteo Marsili、Ilya Nemenman、Alex Nugent、Norman Packard、Dana Randall、Peter Sadowski、Narayana Santhanam、Robert Shaw、Adam Stieg、Elan Stopnitzky、Christof Teuscher、Chris Watkins、David Wolpert、Joshua Yang 和 Yan Yufik。热力学计算。2019年。[10] Jamie Davies 和 Michael Levin。通过主动性与智能体物质实现合成形态学。2022年6月。[11] Thomas Doctor、Olaf Witkowski、Elizaveta Solomonova、Bill Duane 和 Michael Levin。生物学、佛教与人工智能：关怀是智能的驱动力。熵，24(5):710，2022年5月。[12] Chris Fields、Karl Friston、James F. Glazebrook 和 Michael Levin。适用于一般量子系统的自由能原理。生物物理学与分子生物学进展，173:36–59，2022年9月。[13] Chris Fields 和 Michael Levin。作为生命起源与人工生命研究模型的调节性发育。2022年10月。[14] Steven A Frank。生物系统中的输入—输出关系：测量、信息与希尔方程。生物学直报，8(1)，2013年12月。[15] Karl Friston、Lancelot Da Costa、Noor Sajid、Conor Heins、Kai Ueltzhöffer、Grigorios A. Pavliotis 和 Thomas Parr。自由能原理：更简单但又不至于过于简单。2022年。[16] Mandi Gandelman、Mark Levy、Patricia Cassina、Luis Barbeito 和 Joseph S. Beckman。P2x7 受体通过依赖过氧亚硝酸盐/FAS 的通路诱导运动神经元死亡。神经化学杂志，126(3):382–388，2013年6月。[17] Stephen Grossberg。竞争、决策与共识。数学分析与应用杂志，66(2):470–493，1978年11月。[18] Stephen Grossberg。生物学竞争：决策规则、模式形成与振荡。美国国家科学院院刊，77(4):2338–2342，1980年4月。[19] Stephen Grossberg。行为经济学与神经经济学：合作、竞争、偏好与决策。载于*2013*年国际神经网络联合会议（*IJCNN*）。IEEE，2013年8月。[20] Julia J. Harris、Renaud Jolivet 和 David Attwell。突触能量的消耗与供应。神经元，75(5):762–777，2012年9月。

[21] Donald D. Hoffman、Manish Singh 和 Chetan Prakash。《感知的界面理论》。Psychonomic Bulletin I& Review，22(6):1480–1506，2015年9月。[22] Jakob Hohwy。《自我证据化的大脑》。Noûs，50(2):259–285，2014年3月。[23] Jacek Izydorczyk。《摩尔定律热噪声终结的三个步骤》。IEEE Transactions on Very Large Scale Integration (VLSI) Systems，18(1):161–165，2010年1月。[24] Kuno Kim、Shivam Garg、Kirankumar Shiragur 和 Stefano Ermon。《逆强化学习中的奖励识别》。139:5496–5505，2021年7月18–24日。[25] Michael D. Kirchhoff 和 Tom Froese。《有生命之处便有心智：支持强生命—心智连续性论题》。Entropy，19(4):169，2017年4月。[26] Laszlo B Kish。《摩尔定律的终结：微纳电子集成的热（噪声）终结》。Physics Letters A，305(3-4):144–149，2002年12月。[27] Joel Lehman、Jonathan Gordon、Shawn Jain、Kamal Ndousse、Cathy Yeh 和 Kenneth O. Stanley。《通过大型模型进行进化》。2022年。[28] Joel Lehman 和 Kenneth O. Stanley。《放弃目标：仅通过寻找新颖性实现进化》。Evolutionary Computation，19(2):189–223，2011年6月。[29] Michael Levin。《“自我”的计算边界：发育生物电驱动多细胞性与无尺度认知》。Frontiers in Psychology，10，2019年12月。[30] Michael Levin。《普遍心智的技术路径：一种基于实验的框架，用于理解多样的身体与心智》。Frontiers in Systems Neuroscience，16，2022年3月。[31] Michael Levin 和 Christopher J. Martyniuk。《生物电密码：一种用于动态控制生长与形态的古老计算媒介》。Biosystems，164:76–93，2018年2月。[32] Warren Mansell，编。《知觉控制理论跨学科手册：知觉控制理论跨学科手册》。Academic Press，美国加利福尼亚州圣迭戈，2020年5月。[33] Richard Marken、Richard Kennaway 和 Tauseef Gulrez。《行为错觉：snark 是 boojum》。Theory I& Psychology，32(3):491–514，2022年1月。[34] Richard S. Marken 和 Dennis M. Shaffer。《运动的幂律：行为错觉的一个实例》。Experimental Brain Research，235(6):1835–1842，2017年3月。[35] Stuart M. Marshall、Cole Mathis、Emma Carrick、Graham Keenan、Geoffrey J. T. Cooper、Heather Graham、Matthew Craven、Piotr S. Gromski、Douglas G. Moore、Sara. I. Walker 和 Leroy Cronin。《利用组装理论和质谱法将分子识别为生物特征》。Nature Communications，12(1)，2021年5月。[36] Stuart M. Marshall、Douglas G. Moore、Alastair R. G. Murray、Sara I. Walker 和 Leroy Cronin。《利用组装空间形式化生命的演化路径》。Entropy，24(7):884，2022年6月。[37] Stuart M. Marshall、Douglas G. Moore、Alastair R. G. Murray、Sara I. Walker 和 Leroy Cronin。《利用组装空间形式化生命的演化路径》。Entropy，24(7):884，2022年6月。[38] T. H. Morgan。《斑纹涡虫再生的实验研究》。Archiv für Entwicklungsmechanik der Organismen，7(2-3):364–397，1898年10月。[39] Jeremy E Niven。《神经元能量消耗：生物物理学、效率与进化》。Current Opinion in Neurobiology，41:129–135，2016年12月。[40] Graham Oddie。《Armstrong论爱利亚原则与抽象实体》。Philosophical Studies，41(2):285–295，1982年3月。[41] Vaibhav P. Pai、Sherry Aw、Tal Shomrat、Joan M. Lemire 和 Michael Levin。《跨膜电压电位控制非洲爪蟾胚胎眼部图式形成》。Development，139(2):313–323，2012年1月。[42] Mauricio H. Pontes、Anastasia Sevostyanova 和 Eduardo A. Groisman。《ATP过多为何不利于蛋白质合成》。Journal of Molecular Biology，427(16):2586–2594，2015年8月。

[43] William T. Powers。《目的性系统的定量分析：科学心理学基础的一些初步工作》。心理学评论，85(5):417–435，1978年9月。[44] William T Powers。《行为：知觉的控制》。Benchmark Publications，第2版，2005年3月。[45] Michiel W. H. Remme、John Rinzel 和 Susanne Schreiber。《功能与能量消耗对典型计算中的神经元生物物理学的约束：巧合检测》。*PLOS*计算生物学，14(12):e1006612，2018年12月。[46] E.D. Schneider 和 J.J. Kay。《生命作为热力学第二定律的体现》。数学与计算机建模，19(6-8):25–48，1994年3月。[47] Erwin Schrodinger。《什么是生命？：兼收《心智与物质》及自传性随笔》。2013年6月。[48] Abhishek Sharma、Dániel Czégel、Michael Lachmann、Christopher P. Kempes、Sara I. Walker 和 Leroy Cronin。《组装理论解释并量化选择与进化的涌现》。2022年。[49] Lakshwin Shreesha 和 Michael Levin。《发育层的能力改变形态发生人工胚胎模型中的进化动力学》，2022年。[50] L. Soros 和 Kenneth Stanley。《通过 Chromaria 人工生命世界确定开放式进化的必要条件》。ALIFE 14：第十四届生命系统合成与模拟国际会议：793–800，2014年7月。[51] N. Stepp 和 M.T. Turvey。《论强预期》。认知系统研究，11(2):148–164，2010年6月。[52] Richard C Tolman。《相对论、热力学与宇宙学》。Dover物理学丛书。Dover Publications，美国纽约州米尼奥拉，2003年3月。[53] Jan Toman 和 Jaroslav Flegr。《基于稳定性的排序：被遗忘的过程，亦是（不仅仅是）生物进化背后的过程》。理论生物学杂志，435:29–41，2017年12月。[54] Henry Veatch。《科学解释的若干方面及科学哲学其他论文》。科学哲学，37(2):312–314，1970年6月。[55] Rui Wang、Joel Lehman、Aditya Rawal、Jiale Zhi、Yulun Li、Jeff Clune 和 Kenneth O. Stanley。《增强型 POET：通过无限创造学习挑战及其解决方案实现开放式强化学习》。2020年。[56] Stephen Wolfram。《一类可能用于表征基础物理学的模型》。2020年。[57] M V Zamaraeva、R Z Sabirov、E Maeno、Y Ando-Akatsuka、S V Bessonova 和 Y Okada。《细胞凋亡过程中细胞质ATP升高时细胞死亡：利用细胞内荧光素酶开展的生物发光研究》。细胞死亡与分化，12(11):1390–1397，2005年5月。

术语表

- **组装理论** ——一种生命物理理论，通过测量并应用递归算法，对生命进行分类和检测。该理论将不同的生命形式置于复杂性谱上。

- **可说服性轴** ——一种观点，认为每个“事物”在控制它所需的努力程度上都处于一个连续谱上。所需努力与“事物”认知光锥的复杂性／大小成反比。

- **认知** ——本文将其操作性地定义为某个“事物”能够“感知”“相信”和“行动”的程度。换言之，与之相关的复杂性
- **认知能力测试** ——指用于确定任何“事物”认知光锥的一组测试。其中一些测试包括：受控变量测试（TCV）、变量可控性测试（TVC）和控制开关测试（TCS）。

- **认知光锥** ——描述某个事件在时空中距离多远时，仍能影响某个系统正在进行的计算。

- **能量稳态循环** -“能量稳态回路”的简称，是一种控制系统的抽象：它通过改变自身结构，学习调节自身的“能量水平”或特定动力学。它对应于自适应热力学开放系统或自适应耗散结构的概念。能量稳态循环代表结构之间的“关系”。

- **Ensoul** -“通过能量稳态网络实现自组织智能超低功耗系统”（Enerstatic Network approach to Self Organizing intelligent Ultra Low Power Systems）的缩写。本文提出这一框架，以便更好地理解现实的本质。
- **环境能量稳态回路（EEL）**-“环境能量稳态回路”的简称，是一种用于实现能量封闭系统的抽象。EEL 负责创建结构和能量稳态循环。

- **反馈** -事物的作用或“输出”对其成因或“输入”的影响程度。
- **框架** -框架是一套规则、思想或信念，人们借此解决问题或作出决策。

- **自由能原理**- 对任何自组织系统必须具备的属性进行数学形式化。

- **目标** - 某事物的行为（即因果力）所期望达成的结果。
- **智能** - 本文将其操作性地定义为某个“事物”实现其存在“目标”的适应能力。

- **生命—心智连续性论题** - 一种哲学观点，认为生命与心智同时出现，此后以日益复杂的方式显现。

- **新心理学家** - “对 SOULS 的新研究”——对运用本文所述框架和方法的人可能采用的称谓。具体而言，指主要致力于研究任何“事物”的可测量的基质不变且操作性定义的属性（即智能、认知、压力水平等）的人。

- **开放式演化** - 在此过程中，新的“事物”会在看似无限期的时间内不断产生。这一过程被认为体现于我们的宇宙之中，且组装理论对其作出了很好的描述。

- **泛计算主义** - 一种形而上学观念，认为存在的每一个“事物”都可以被理解为正在“计算”某种东西。

- **自我** - 一组共享不同程度应力的部分
- **自我证据化** - “self-evident”的动词形式；如果某个解释所解释的事件为该解释提供了关键证据，那么该解释就是自我证据化的。任何“事物”的行为都会通过存在性反馈回路，为其自身的存在提供关键证据。

- **SOULS** - “自组织智能超低功耗系统”的首字母缩写，旨在概括 Ensoul 致力于理解和利用的“事物”所具有的多尺度基底无关性质，以创造新颖技术。

- **应力** - 任何将某个“事物”推离其“偏好”或“健康”状态的作用。
- **实体一元论** - 一种形而上学观念，认为每个“事物”都由同一种“物质”构成。
- **IOS** - “输入输出系统”的缩写，用于概括其“波动”或目标导向行为已被抑制的 SOULS。

- **TAME** - “普遍心智的技术路径”的缩写，是 Michael Levin 开发的一种用于理解多样身体与心智的框架。

- **目的论** - 具有目的性或目标导向的行为。
- **目的拟态** - 并非真正具有目标导向性，但看起来“仿佛”具有目标导向性的行为。
- **冯·诺依曼架构** - 一种基于 John von Neumann 于 1945 年所作描述的计算机架构。这是如今大多数计算机所采用的架构。其含义已演变为任何“存储程序计算机”，其中指令取指和数据操作无法同时进行。