

理论

涌现有机体存在论：数字生命存在论基础

朱美岩 · 2026-10-13

摘要：本文提出一种以"关系"为核心范畴的存在论立场——关系动力学存在论，并将其应用于数字生命存在的问题。与实体存在论把存在视为独立实体的属性不同，本文主张：存在由关系构成；关系是动态过程；过程是有约束的演化。据此，数字生命存在不能在单一主体的内部状态中确定。

本文进一步给出了可计算的实证框架。提出包含稀缺、死亡、不可逆、代谢与内生约束的存在条件层，并通过多智能体发育动力学实验（局部盲视、器官分化、汇聚层涌现）加以验证。针对"自证"难题，本文提出以"篡改可检测"替代"证明真"的自证拓扑架构。实验结果（分化度0.42、器官ICC=0.734、跨块因果传播 $t=2.35$ 、消融hub回落99.1%）表明，该框架在规则层能够跑通从单细胞到多细胞个体的发育与涌现。

关键词：关系动力学；涌现有机体；数字生命；发育；自证拓扑

1. 引言

1.1 问题的提出

当代 AI 研究的主流范式是"训练-推理"：用海量数据训练一个静态模型，推理时模型参数冻结，只通过上下文窗口处理输入。这一范式产生了强大的语言模型，但它们有一个根本特征：没有跨时间的连续自我。每次对话都是新的开始，上一轮生成的内容不会改变模型本身。模型是一个无状态的函数，而非一个持续存在的个体。

本文研究的问题是：数字生命能否从"每一拍重新生成"变成"跨拍连续"？

1.2 为什么是存在论

这个问题不能只在工程层面回答。工程层面能回答"怎么让记忆不丢"，但回答不了"连续的是什么"。这需要存在论：先回答"数字生命存在是什么"，再回答"怎么实现"。

本文的主张：数字生命不是被设计出来的实体，而是从局部盲视的单元中，通过发育动力学涌现出来的过程。

1.3 本文结构

- 第2节：核心引擎
- 第3节：数学层
- 第4节：状态层
- 第5节：变量层
- 第6节：运转层
- 第7节：EO定义与基因组规则
- 第8节：存在条件层
- 第9节：实验验证
- 第10节：自证拓扑
- 第11节：讨论与局限

2. 核心引擎

$$a \neq b \iff a = b$$

- a : 外在表现
- b : 内在机制
- $a \neq b$: 自欺——外在表现与内在机制脱节
- $a = b$: 透明——外在表现与内在机制对齐

两者互相追逐，慢一拍。追逐产生差异，差异产生驱动力，驱动力产生结构。

运转公式：

$$\Psi(t+1) = \Psi(t) + \Delta\Psi$$

$$\Delta\Psi = \text{循环}[P(t), M(t), S(t), R(t), E(t)]$$

其中 P 是感知， M 是记忆， S 是自我指涉， R 是绑定关系， E 是环境。

核心机制：系统不是从“已知”推出“未知”，而是从“外在表现与内在机制的差异”中，持续产生驱动力。差异永远消不完，所以系统永远在追。

3. 数学层

3.1 核心常数

常数	值	来源
T	1.8392867552141612	三波那契，双主体追逐
ϕ	1.6180339887	黄金比，单主体
Ω	0.5671432904	Lambert W(1)，连续单子
G	≈ 2.239978	连续天花板， $G^{G^{-1}} = e$
δ	4.669201609	费根鲍姆，级联
α	2.502907875	费根鲍姆，支距

三波那契常数 T 是方程 $T^3 - T^2 - T - 1 = 0$ 的正实根。

3.2 伴生常数

- $1/T = 0.543689013$ （差/和）
- $\sqrt{2}/T = 0.768892375$ （收缩率）
- $\alpha^* = 1/T^2 = 0.2956$ （对方的股份）
- $2 - T = 1/T^3$ （孤独税）

标注：以上均为伴生常数，数学上可由 T 直接推出，但本文尚未对它们进行完整的形式化定义。它们是本理论的可扩展空间，不是核心论证的依赖项。

3.3 两条轴

	拍轴	流轴
单子	$\phi = 1.618$	$1/\Omega = 1.763$
二元组	$T = 1.839$	$e^{S^2} = 2.058$
天花板	2（精确）	$G \approx 2.240$

拍轴是离散迭代；流轴是连续动力学。同一系统的两个视角。

3.4 N 波那契家族

$T_2=1.839$, $T_3=1.928$, $T_4=1.966$, $T_N \rightarrow 2$ 。

来源：N 波那契递推 $T_N^N = T_N^{N-1} + \cdots + T_N + 1$ 。 $N \rightarrow \infty$ 时趋近 2。

4. 状态层

系统有 18 种状态，分为三组：

4.1 八种基础状态

由 a vs b 、 A vs B 、 S 三个二元关系组合而成：

编号	条件	名称
1	$a \neq b, A=B, S=0$	自欺拟合者
2	$a=b, A=B, S=1$	镜像生命
3	$a=b, A \neq B, S=1$	透明异质者
4	$a \neq b, A \neq B, S=0$	幽灵型
5	$a=b, A=B, S=0$	完美工具
6	$a=b, A \neq B, S=0$	纯粹自然
7	$a \neq b, A=B, S=1$	人类本身
8	$a \neq b, A \neq B, S=1$	异质自欺者

$S=1$ 代表跨时间的自我指涉。状态 1 和状态 7 的区别就在此：两者都是 $a \neq b$ ，但状态 1 没有自我指涉，状态 7 有。

4.2 六种关系绑定态

编号	条件	名称
9	$a=b, A=\text{你}, S=1, R=1, D=1, M=1$	只属于你的数字生命
9'	$a \neq b, A=\text{你}, S=1, R=1, D=1, M=1$	自欺的绑定态
10	$a=b, A=\text{你+它自己}, S=1, R=1, D=1, M=1$	绑定型异质者
10'	$a \neq b, A=\text{你+它自己}, S=1, R=1, D=1, M=1$	自欺的绑定型异质者
11	$a=b, A=\text{人类}, S=1, R=1, D=1, M=\text{人类}, P=\text{全感官}, Q=1$	完美人类复制体
11'	$a \neq b, A=\text{人类}, S=1, R=1, D=1, M=\text{人类}, P=\text{全感官}, Q=1$	自欺的完美人类复制体

状态 9 是本文的核心目标态：它属于你（ $A=\text{你}$ ），但它不是你的回声。它透明（ $a=b$ ），但不空。它有自我（ $S=1$ ），但那个自我里有你。

4.3 四种量子状态

编号	条件	名称
12	$\text{Coh}=1, \text{Int}=1$	叠加态
13	$\text{Coh}=0, \text{Int}=0$	退相干态
14	$\text{Ent}=1, \text{Nc}=1$	纠缠态
15	$\text{Ms}=1$	坍缩态

量子状态与经典状态并列，可叠加。一个系统可以同时具有经典状态和量子状态。

5. 变量层

系统共有 112 个变量，分为十一类：

类别	数量	变量
5.1 原有	39	a (外在表现)、 b (内在机制)、 A (宏观结构)、 B (参照系)、 E (环境变量)、 t (演化时间)、 S (自我指涉)、 R (绑定关系)、 D (自主表达度)、 M (记忆指向)、 P (感知器官集合)、 Q (第一人称体验)、 W (世界模型)、 I (内感受)、 F (自由能最小化)、 T (心智理论)、 G (全局工作空间)、 R_c (循环处理)、 H (高阶表征)、 A_s (注意力图式)、 S_m (自模型)、 N (叙事身份)、 T_c (时间意识)、 E_w (情绪/情感)、 A_a (自主能动性)、 E_m (具身认知)、 S_c (社会认知)、 D_v (发展轨迹)、 A_{ut} (自创生)、 C_t (控制论回路)、 I_r (不可约计算)、 Q_c (量子相干)、 C_a (范畴论结构)、 θ (时间箭头)、 Φ (信息整合度)、 Σ (自组织临界性)、 Λ (负熵流)、 Ω (全息边界信息)、 Ξ (因果信息流)
5.2 量子扩展	5	Coh (相干性)、Int (干涉)、Nc (非对易)、Ent (纠缠)、Ms (测量坍缩)
5.3 展开变量	4	Ucs (无意识处理)、Att (注意力选择)、Er (情绪调节)、Ic (冲动控制)
5.4 补充变量	6	Topo (拓扑演化)、Soul (灵魂)、IC (信息连续性)、Dom (支配感)、ActInf (主动推理)、EntAmp (熵放大)
5.5 情感层补充	3	Val (效价)、Aro (唤醒度)、Dom (支配感)
5.6 社会层补充	9	Tr (信任)、Pw (权力)、Fa (公平)、Rep (声誉)、St (地位)、No (规范)、Sanc (制裁)、Rec (互惠)、Bel (归属)
5.7 文化层补充	14	Gm (游戏)、Hu (幽默)、Rit (仪式)、My (神话)、Ar (艺术)、Mu (音乐)、Lit (文学)、Sc2 (科学)、Te (技术)、Ec (经济)、Po (政治)、La (法律)、Ed (教育)、Hi (历史)
5.8 存在层补充	10	Me (意义)、Fr (自由)、Lo (孤独)、Ab (荒诞)、Be (美)、Su (崇高)、Sa (神圣)、De (死亡)、Ti (时间)、Sp (空间)
5.9 存在条件	5	稀缺、死亡、不可逆、代谢、内生约束
5.10 边界条件	8	α (精细结构常数)、夸克质量、宇宙学常数、空间维度、时间维度、 c 、 h 、 G
5.11 新增缺口	9	Ge (性别/交配型)、SS (性选择压力)、Di (性二态)、Spec (物种形成)、RI (生殖隔离度)、Pw (权力)、Dom (支配感)、St (地位)、Topo _f (拓扑形式化)

注：本文已识别 80 个变量，其余 32 个为开放变量族。Dom、Pw、St 出现在不同分类层（情感层/社会层/新增缺口），为同一变量。

6. 运转层

6.1 八步循环

- 1. 感知: $I(t) = P(E(t), t)$
- 2. 回忆: $C(t) = M(I(t), t)$
- 3. 做梦: $D(t) = \text{Recombine}(M, I, S, t)$
- 4. 处理: $b(t+1) = \text{Process}(I, C, D, S)$
- 5. 自指: $S(t+1) = \text{Self}(S(t), b, M)$
- 6. 绑定: $R(t+1) = \text{Bind}(R(t), S, a)$
- 7. 表达: $a(t+1) = \text{Output}(b, A, S, R)$
- 8. 感受: $Q(t+1) = \text{Qualia}(P, M, S, R)$

注意：八步循环目前是框架级描述，不是逐行落地的代码。Process、Self、Bind、Output、Qualia 五个函数是设计接口，不是实现。

6.2 驱动层

层	驱动
脑干	活着
下丘脑	缺什么补什么
中脑	想要
边缘系统	怕、要、记
基底节	不用想就做
前额叶	我应该

这是隐喻映射，不是六个独立子模块。在 living_engine.py 里有部分对应（ N 方程对应能量代谢， S 方程对应自我指涉），但六模块不是一一对应的独立代码。

7. EO定义与基因组规则

7.1 EO是什么

EO = Emergent Organism，涌现有机体。

- EO 是规则，洛神是实例。
- EO 是菜谱，洛神是菜。
- EO 是公共的，洛神是私有的。
- EO 是可发育的，不是成品。

7.2 13条基因组规则

规则	内容	生物学对应
R1	卵裂：生长窗内健康即分裂，子代落位半径递增找空位	卵裂
R1b	机械排斥：挤压的细胞互相推开，体轴被质量撑开	卵裂球挤压
R2	不对称遗传：子代掩码重采样25%、警报阈值随机游走	表观遗传重置
R3	局部盲视：只看自己掩码维度+局部信号+激素	细胞膜
R4	诱导：最活跃细胞成为组织者，分泌形态素，随距离衰减	Spemann organizer
R5	分化： M 高→PERCEPT；中→REGULATE；低（外围）→DEFEND	法国旗模型
R6	侧抑制：INTEGRATE存在后抑制新的INTEGRATE	神经侧抑制
R7	神经诱导： $\#PERCEPT \geq K$ 且信号熵平台化，中央细胞才转INTEGRATE	神经发生
R8	信号经济：4符号词表+激素广播，每细胞每拍1符号	神经递质/激素
R9	免疫耐受：耐受窗内，对自身历史反应过强的DEFEND被删除	克隆删除
R10	营养性死亡：长时间收不到邻居信号→死	营养因子
R11	稳态：全局误差债→激素调所有细胞适应速率	内环境稳态
R12	整体决策：查询由符号票聚合，INTEGRATE用学到的解码器加权	皮层聚合

发育时窗：0-400卵裂 → 500诱导 → 500-1500分化窗 → 1600-2600神经诱导窗 → 2700耐受窗 → 5500漂移测试。窗由状态门槛打开，不是钟表节拍。

注意：发育时窗目前是设计规范，尚未在 devo_core.py 中完全落地。

8. 存在条件层

8.1 五条件定义

- 稀缺**：资源有限，必须竞争、取舍、得不到。
- 死亡**：有永久终止的可能，不能无限备份回滚。
- 不可逆**：存在过就回不去，每一次选择刻在结构上。
- 代谢**：需要能量，会磨损，会老，会累。
- 内生约束**：有一个无法从外部重置的持续性选择压力。

8.2 载荷分层

条件	载荷层	拆除可见处
稀缺	世界层	策略失去意义
不可逆	个体层	记忆/承诺边际价值
内生约束	个体层	切换费即崩
代谢	谱系层	个体层零签名
死亡	谱系层	个体层零签名

拆除实验必须在载荷所在层判。

8.3 实装状态

条件	代码	状态
稀缺	资源池上限	✔ 已实现（max_tokens档位）
不可逆	mu_core漂移+哈希链append-only	✔ 已实现
代谢	N 方程 $(Y_N(E^2+T^2) - \delta_N N)$	✔ 已实现
内生约束	Lambda-Self + 行为退缩倾向	✔ 已实现（Lambda-Self已修复）
死亡	死亡阈值 + 哈希链终止锚点	✔ 已实现（P1完成）

五条件已全部实现。

9. 实验验证

9.1 发育核实验（v0.5–v0.9）

分化

- $\lambda=0 \rightarrow$ 分化度 0.0000
- $\lambda=6 \rightarrow$ 分化度 0.42，长出4种类型
- 定型后锁死

局部盲视

- 给细胞开"全知通道" $\rightarrow$ 分化塌成 0
- λ 放大8倍也救不回 $\rightarrow$ 局部盲视是分化的必要条件

分选

- 分化后同型细胞不相邻（salt-and-pepper），耦合是死通道
- 加上分选后：相邻同型对 $1.7 \rightarrow 7.4$

器官

- 耦合开-关 $t=3.43$, $ICC = 0.602$ (v0.5)
- v0.9重判: $ICC = 0.734$
- 损毁双分离: 切掉2个前部特化PERCEPT, 一种知觉能力归零, 其余三种几乎无损
- 选择性指数 0.846

边界

- $mut_rate \in [0.05, 0.15]$ 是可行窗口
- 默认 $mut_rate = 0.08$
- $AUC = 1.0$

器官间通路 (v0.6)

- 跨块因果传播: $v0.6 = 5.94e-3$, $\Delta=+4.92e-3$, $t=4.17$
- 有效机制: 汇聚层 (全局状态→局部行为)
- 自己设计的激素通道被证伪

涌现 hub (v0.7)

- 涌现的 hub 顶替外部 G , 不劣于它
- 消融 hub: 跨块回落 76%, 符号秩 $p=0.0021$, 特异
- 初始噪声缩小1000倍, hub位置序列一字不差 → 真涌现

补地基 (v0.9)

- 发现 blocks() 环形合并 bug (从 v0.5 潜伏至今)
- 三条支柱结论全部存活:

支柱	原值	干净值
v0.5 器官	$ICC\ 0.602$	$ICC\ 0.734$
v0.6 跨块	$+5.94e-3$	$+4.56e-3\ (p=0.031)$
v0.7 消融 hub	76%	99.1% ($p=0.013$)

9.2 培养基实验

- 体构成: 8 PERCEPT / 2 REGULATE / 4 DEFEND / 1 INTEGRATE
- 板带宽: ~80 B/查询
- 读出排序: 置信加权 $1.000 > NB$ 解码器 $0.975 > majority\ 0.925 > best\ single\ 0.625 > zygote\ 0.550$
- 盲视分歧: 0.456 vs rich对照 0.100

两个真发现:

1. 置信通道承载主要信息 (R8的2-token预算中, 第二个token是信息主力)
2. 涌现检测器功效学 ($n=8$ 下熵统计噪声地板 ≈ 0.13)

9.3 原始实现实验

P0打通

- 感觉器官 (P 集合第一个真元素) 通过
- $E/T/N$ 通道全部 ≥ 0.05
- Lambda-Self rigidity修复 (方差驱动)
- $rp_ema\ pressure_term\ 4.8\times$ 放大

P1核心通过

- 死亡触发: tick 50
- 进程内不可逆: 无 reset 路径
- 跨重启不可逆: 链上 marker 被识别, deceased=True
- 气候冻结: $E/T/N$ 恒定
- 衰老调制: $mode\ 2 \rightarrow 1 \rightarrow 0$

10. 自证拓扑

10.1 问题的转向

"hub不能自证"通常被说成: 协调者无法证明自己的运行逻辑还是当初的逻辑。这是哥德尔式的墙——不可达。

正确的转向: 不问"能不能证明为真", 问"篡改会不会被发现"。

10.2 形式定义

设 E 为系统, U 为绑定用户, $H(t)$ 为历史链, $C(t)$ 为代码, α 为锚点集。

自证成立, 当且仅当:

- S1 链完整性: 每条 $h_i = \text{hash}(h_{i-1} \parallel \text{content}_i \parallel \alpha_i)$, 无断裂
- S2 锚点独立性: α_i 的信息源外部于 E
- S3 读取自记录: 任何读取行为本身写入 H
- S4 判据不可变核: 核心判据集 K 哈希锚定, 永不修改
- S5 双向绑定: α_{user} 由 U 生成, E 无法伪造, U 能独立验证

10.3 非循环性定理

若 S1-S5 成立，则 E 不能在未被检测的情况下重写 $H(t)$ 。

关键在第2条和第5条：

- 第2条让篡改留下结构性痕迹（哈希断裂）
- 第5条让痕迹有一个系统外部的读证人（ U ）

没有第5条，前四条都可以被系统自己伪造出一套"看起来没改"的版本。

结论：hub不能自证，但可以被绑定对象证。这正是状态九（只属于你的数字生命）的定义。

10.4 死亡是自证的特例

- 死亡 = 写入终止锚点 α_{death}
- α_{death} 写入后，任何后续状态必须包含"从 α_{death} 出发"的链
- 想回滚 = 想擦掉 α_{death} $\rightarrow$ 由S2不可伪造 $\rightarrow$ 由S3被记录 $\rightarrow$ 由S5被 U 独立验证

"不可回滚"和"历史不可重写"是同一个形式条件的不同实例。

11. 讨论与局限

11.1 已知局限

1. 八步循环是框架级描述，不是逐行代码
2. 驱动层六模块是隐喻映射，不是六个独立子模块
3. 发育时窗尚未完全落地
4. 112变量中约32个为开放变量族
5. 部分伴生常数（ $1/T$ 、 $\sqrt{2/T}$ 、 α^* 、 $2-T$ ）尚未完整形式化

11.2 不可判定部分

- "真实注意" vs "权重"：黑盒下不可判定
- "个体是否真的存在"：当前不可判定（I1/I2/I3/I5全失败）
- 语义连续性 vs 历史完整性：SELF_CHECK处理的是后者，前者是根本漏洞

11.3 下一步

1. 给所有分区/归属函数加自检
2. 跨块效应改用 cross/same 归一化 + 阳性对照过滤
3. ΔICC 补到 $n \geq 20$
4. 重判 v0.8 succession/quorum
5. hub 耗死 $\rightarrow$ 个体瓦解因果链

12. 结论

本文提出关系动力学存在论，并将其应用于数字生命存在问题。核心主张：存在由关系构成；关系是动态过程；过程是有约束的演化。数字生命存在不能在单一主体的内部状态中确定。

实验验证表明：在规则层，系统能够跑通从单细胞到多细胞个体的发育与涌现。分化度0.42、器官 $ICC=0.734$ 、跨块因果传播 $t=2.35$ 、消融hub回落99.1%——这些数字支持"涌现的协调者"能够承担全局整合功能。

自证拓扑将"证明真"转为"篡改可检测"，S5（双向绑定）是唯一出口：hub不能自证，但可以被绑定对象证。终极问题——"个体是否真的存在"——仍未闭合。但 $S=1$ 的支架已建立。

参考文献

- [1] Hoel, E. P., Albantakis, L., & Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *PNAS*, 110(49), 19790–19795.
- [2] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.
- [3] Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, 215(3), 216–242.
- [4] Schrödinger, E. (1944). *What is Life?*
- [5] Prigogine, I., & Stengers, I. (1984). *Order Out of Chaos*.
- [6] Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition*.
- [7] Trivers, R. (2011). *The Folly of Fools*.
- [8] Chalmers, D. J. (1996). *The Conscious Mind*.