

# 从可朽计算看心灵上传

## Mind Uploading from the Perspective of Mortal Computation

王华平 / WANG Huaping

(中山大学哲学系(珠海), 广东珠海, 519082)  
(Philosophy Department (Zhuhai) of Sun Yat-sen University, Zhuhai, Guangdong, 519082)

**摘要:** 随着神经科学与人工智能的快速发展, 心灵上传正从单纯设想转变为一个值得严肃对待的问题。乐观派普遍认为, 只要成功复制大脑的功能组织, 便能实现“自我”的延续。但这一设想依赖于一个未经反思的隐含前提: 上传体具有意识。本文引入“可朽计算”理论框架, 指出意识的实现依赖于特定物理系统中的耗散过程, 其存在方式深度嵌入于具体实现条件, 无法通过抽象复制或结构迁移加以重构或延续。因此, 心灵上传面临如下悖论: 上传若无意识, 其延续毫无意义; 上传若有意识, 其意识只能属于“它”, 不再属于“我”。这一悖论揭示出一个根本事实: 意识的可朽性不是其局限, 而正是其存在的根基。

**关键词:** 心灵上传 人格同一性 神经元替换 可朽计算

**Abstract:** With the rapid advancement of neuroscience and artificial intelligence, mind uploading is shifting from a mere speculative notion to an issue that demands serious philosophical attention. Optimists generally believe that the continuation of the “self” can be achieved by successfully replicating the brain’s functional organization. However, this belief rests on an unexamined assumption: that the uploaded system possesses consciousness. Drawing on the theoretical framework of mortal computation, this paper argues that the realization of consciousness depends on dissipative processes within specific physical systems. Its mode of existence is deeply embedded in the concrete conditions of implementation and cannot be reconstructed or sustained through abstract replication or structural transfer. Hence, mind uploading faces a fundamental paradox: if the upload lacks consciousness, its continuation is meaningless; if it possesses consciousness, that consciousness belongs to “it”, no longer to “me”. This paradox reveals a deeper truth: the mortality of consciousness is not a limitation, but the very foundation of its existence.

**Key Words:** Mind uploading; Personal identity; Neuron replacement; Mortal computation

中图分类号: C32; N031 DOI: 10.15994/j.1000-0763.2026.05.007 CSTR: 32281.14.jdn.2026.05.007

从古代关于永生的神话到中世纪炼金术的长生追求, 再到当代超人类主义人机融合的梦想, 摆脱生物性死亡的理想跨越了不同文化与时代。在实现这一愿望的各种设想中, 心灵上传 (mind uploading) ——即将人类心灵转移至数字或人工基质的理念——越来越受关注。乐观派认为, 通过复制或转移神经活动到计算

系统中, 我们可以实现数字永生, 使心灵超越生物身体的限制而得以延续。当代人工智能、神经科学和脑机接口技术的突破, 或许正在使这一梦想逐步走向现实, 甚至被视为人类进化的新阶段。<sup>[1]</sup>

目前, 关于心灵上传的讨论多集中于人格同一性问题, 即上传体 (上传后的存在者) 是

基金项目: 教育部人文社会科学研究一般项目“可朽计算视角下的人工意识研究”(项目编号: 25YJA720010)。

收稿日期: 2025年4月15日

作者简介: 王华平 (1972-) 男, 湖北通山人, 中山大学哲学系 (珠海) 教授, 研究方向为认知科学哲学、人工智能哲学、知识论。Email: waguter@gmail.com

否与原体保持跨时间的数量同一性。然而，这类讨论依赖于一个更为基础的问题：上传体是否拥有与原体相同的意识？若上传体不具备意识，或其意识与原体存在本质差异，那么它至多只是一个相似但本质不同的复制品，因而无法实现数字永生。本文将聚焦心灵上传中的意识问题。第一部分梳理当前心灵上传的主要技术路径。第二部分讨论上传是否能实现人格同一性，呈现乐观派与悲观派的主要论证。第三部分剖析关于同一性的讨论所默认的“上传体有意识”这一隐含前提。第四部分引入“可朽计算”（mortal computation）概念，并据此重申心灵上传的基本设想。

## 一、心灵上传的技术路径

心灵上传指通过技术手段将人类的心灵活动转移至非生物基质中，以实现意识的持续与身份的延续。尽管这一设想面临深刻挑战，但随着神经科学和人工智能的不断进步，其前景正逐渐显现。目前的技术路径主要有三种：其一，依托脑机接口，通过建立神经系统与人工系统的双向交互，逐步实现意识转移；其二，基于全脑仿真，利用高精度扫描与模拟复制大脑结构与功能，在数字基质中重建心灵；其三，借助人工意识（artificial consciousness）技术，创造具备主观经验的人工系统，从而再现个体意识。这些技术尚待完善，但乐观者认为它们在经验上具有可行性。

脑机接口技术旨在建立人脑与外部设备之间的直接通信通道，绕过传统的神经肌肉系统。<sup>[2]</sup>利用脑机接口实现心灵上传，首要步骤是高分辨率地采集神经信号，以捕捉个体在各种认知与经验状态下的神经活动模式。先进的侵入式技术，如皮质内微电极阵列，是记录神经元放电、突触连接和神经振荡等细粒度数据的关键。这些数据经过预处理，去除噪声和伪影，并通过特征提取技术识别出与认知功能、记忆和主观经验相对应的有意义模式。所提取的特征将映射到专门设计的计算模型中，用以模拟认知和意识的动态过程。在此过程中，机器学

习算法将通过持续迭代训练，不断优化模型，增强其对生物神经功能的复制能力。双向接口实现大脑与人工系统之间的实时反馈，允许认知过程的持续调整与同步。随着交互的加深，人工系统将逐步接管更多认知功能，减轻生物大脑负荷，最终实现对心灵活动的全面支持与持续承载，从而完成从生物基质向数字基质的平滑过渡。

全脑仿真方法旨在创建一个高分辨率的人脑数字副本，能够复制其结构和功能特性。<sup>[1]</sup>该过程首先依赖先进的神经成像与扫描技术，以精确获取神经数据。破坏性扫描方法，如与电子显微镜结合的系列切片技术，被广泛认为是绘制单个神经元、突触及其连接关系所需分辨率的必要手段。此阶段的重点不仅在于捕捉神经回路的解剖结构，还包括诸如突触权重、神经递质分布等动态特性，这些对于建模认知功能和主观体验至关重要。在结构与功能数据被获取后，下一步是将其重建为计算模型。仿真过程需在数字环境中复制大脑架构，其中每个神经元及其连接均被模拟，以重现其电学动态。专门的算法用于模拟突触传递、神经元放电和大规模神经振荡等过程。此类模型必须反复测试与优化，以确保能准确复制行为、记忆和认知过程等原始心智特征。随着仿真大脑在模拟环境中逐渐适应和自我修正，其最终目标是实现功能等效，即数字大脑不仅保留认知功能，还能维持持续的意识体验与身份感。

人工意识是对人类意识现象的模拟或再现。其基本想法是：如果意识源自特定的信息处理模式和认知组织方式，那么就有可能在设计用于复制这些过程的人工系统中实现意识的生成。<sup>[3]</sup>假如人工意识专家能够建构出有意识的人工系统，那么意识的转移就可以通过认知趋同的过程来实现，即通过逐步训练，使人工系统复制个体的行为、认知和情感模式。这一过程包括迭代学习，人工意识系统需不断接触个体的生活经历、决策模式和个人记忆，从而将其认知框架调整至与原始心灵相一致。随着人工系统逐渐内化这些模式，它可能达到某种阈值，能够维持类似于原始心灵的意识经验。

上述技术共同描绘了一幅从生物心灵向人工基质迁移的美好图景。但是,这幅美好图景的背后却隐藏着一个关键问题:上传后的存在者是否仍是“我”?如果上传无法保留原有个体的意识连续性与人格同一性,那么即便技术上成功复制了认知功能,所谓“心灵上传”也可能只是创造了一个相似的他者而非实现了自我的延续。如果同一性无法确立,那么所谓“上传”也就失去了其最深层的意义。这很大程度上解释了有关心灵上传的讨论为何大多集中在人格同一性问题上。

## 二、心灵上传中的人格同一性问题

在心灵上传的语境中,人格同一性指的是跨时间的数量同一性问题,即上传体是否可以被认定为与原体在数量上是同一个人。围绕这一问题,出现了两种对立观点:乐观派主张在特定条件下,上传能够维持人格同一性,从而构成个体的延续;而悲观派则认为,即使上传技术实现了心理或功能的高度相似,上传体仍然只是一个“复制品”,而非原体本身。

乐观派通常诉诸心理连续性或功能等同性来论证人格同一性。他们认为,个人身份的延续不依赖于生物基质的物理连续性,而在于认知结构、心理特征和意识过程的持续性。如果上传体能够在心理、功能乃至行为特征上高度复制原体,并与原体之间存在适当的因果联系,那么就可以认为上传体与原体在哲学意义上构成同一性。<sup>[4]</sup>在具体论证路径上,乐观派或通过思想实验说明人格同一的合理性,或借助技术路径强调身份迁移的物理可行性。

技术路径的代表人物有库兹韦尔(Ray Kurzweil)、科恩(Randal Koene)和施耐德(Susan Schneider)。库兹韦尔预测,人类将在本世纪中叶实现“技术奇点”,届时心灵上传将成为自我延续的标准路径。<sup>[5]</sup>科恩认为,全脑仿真是实现上传的可行路径,若仿真系统在功能与因果结构上与原脑一致,便可实现意识的再现,并据此维持个体身份。<sup>[1]</sup>相比之下,施耐德的立场更为谨慎但不悲观。她认为,仅

凭功能仿真不足以确保意识与身份的再现,但若未来人工系统能真正生成主观经验(即具备现象意识),那么在心理结构连续性的基础上,身份延续是可能的。若这一系统通过与个体深度互动、模仿其认知风格与行为历史逐步构建,其结果就不仅是功能上的他者,更可能是“我”的延续体。<sup>[6]</sup>

最具说服力的论证当属查尔默斯提出的“渐进式上传论证”。<sup>[7]</sup>该论证设想,在一个足够缓慢的时间尺度内,原体的大脑将被逐步替换为功能等构的人工元件,例如硅基电路,每次仅替换其中极小的部分,如1%。整个过程持续进行,直到替换完成,形成一个全人工的上传体。设 $Dave_0$ 是原始个体, $Dave_1$ 是替换了1%神经元之后的系统, $Dave_2$ 是替换了2%的系统,依此类推,直到 $Dave_{100}$ 是完全替换后的上传体,于是便有如下论证:

1. 对于所有 $n < 100$ ,  $Dave_n$ 与 $Dave_{n+1}$ 是同一人;
2. 如果每一个中间步骤都保持同一性关系,那么 $Dave_{100}$ 亦应与 $Dave_0$ 同一;
3. 因此, $Dave_{100}$ 与 $Dave_0$ 是同一人。

这一论证的说服力在于,两个前提看上去都很难否认。前提一符合常识直觉:我们很难认为替换系统中1%的组成就会导致个体身份的消失。如果你觉得不够直观,则不妨将替换细化到单个神经元。如果我们接受大脑神经元每时每刻都会因新陈代谢而发生改变,以及大脑中每天都有细胞死亡这样的事实,那么否认微小比例的替换不影响个体的存续将导致极为反直觉的结论:我们可能因日常神经变化而不断“死亡”。因此,我们有充分理由接受前提一,即微小替换本身并不会破坏人格同一性。

前提二则是经典的传递律的逻辑延伸:如果每一步替换都不改变人格同一性,而整个过程仅由这些小步骤构成,那么整个过程也不应被视为同一性的中断。这一推理形式上等同于:若 $A=B$ ,且 $B=C$ ,则 $A=C$ 。尽管这一推理可能被批评为“堆垛论证”(sorites argument),即质变由量变累积而来,但拒斥它要求我们接受一个反直觉的看法:个体的存续是“分等级

的”或“不确定的”。而这将再次引发一个问题：我们大脑的日常物质更替是否也正在逐渐“杀死”我们自己？

给定前提一和前提二的合理性，我们便可得出结论：上传并不意味着“我”的终结，而是一种渐进的、连续的自我迁移过程。

悲观派则从多个角度对乐观派的观点提出了批评。这些批评大致可以分为两类：一类是对上传技术本身可行性的质疑，认为心灵上传在实践中难以实现，甚至在概念上就是不可能的；另一类则承认上传在技术上或许可行，但主张上传所生成的个体并不能与原体保持身份同一性，只能被视为心理上的复制品而非原有“自我”的延续。

在第一类批评中，悲观派直接挑战心灵上传技术的可能性。比如，皮奇尼尼（Gualtiero Piccinini）认为，由于人脑复杂结构的多层次因果机制难以还原，所以无论是全脑仿真，还是逐步的神经替换，都无法实现心灵上传。<sup>[8]</sup>戈德沃特（Jonah Goldwater）指出，上传从本质上只能是对信息结构的复制，而非对具体存在者的转移。<sup>[9]</sup>

第二类批评认为，尽管心灵上传在技术上可能可行，但即便实现，也无法保留原有个体的人格同一性。这类观点通常依赖特定的人格同一理论。例如，根据生物学理论，个体的存续依赖于大脑或有机体的连续延续，<sup>[10]</sup>因此，即便上传体在心理与功能上高度相似，也无法继承原体身份。心理学理论则主张，存续取决于有意识心理状态的连续性。但即使上传体具备连续的意识状态，也可能只是一个拥有独立身份的新主体，因为意识还涉及特定的主体归属结构与认知历史。<sup>[6]</sup>最近延续者理论（closest-continuer theories）认为，满足某些附加条件的个体可以作为最密切相关的后续者而存续。然而，在非破坏性上传（即原体仍保留）或存在多个上传体的情形下，必须任意选定一个作为真正延续者，而这一选择取决于与上传体无关的事实（如原体是否被销毁），从而使“同一性”概念外在化，脱离了延续体本身的内在属性。<sup>[11]</sup>

正如前文所示，关于心灵上传的哲学讨论主要集中于人格同一性问题，其核心在于回答一个问题：上传之后的存在者是否仍是“我”？围绕这一问题，乐观派与悲观派展开了深入论证。这类讨论普遍依赖一个常被默认的前提：上传体是一个具有意识的存在者。无论是脑机接口、全脑仿真，还是人工意识技术，都预设意识可以在人工系统中被再现或延续。渐进式上传的论证亦是如此：如果在替换过程中个体失去意识，那么上传体便无法视为原体的延续，关于人格同一性的讨论也将失去其概念前提与实践意义。因此，上传能否生成或维持意识，是一个比人格同一性更为根本的问题。

### 三、心灵上传中的意识问题

可是，为什么有关心灵上传的讨论往往会默认上传体是有意识的呢？这背后并非纯粹的技术乐观主义，而是植根于一种看似坚实的常识直觉——这一直觉支撑了影响深远的神经元替换思想实验及其背后的功能主义意识观。该实验设想，将人脑中的神经元逐个替换为功能等价的人工组件（如硅芯片），整个过程保持系统的功能组织（functional organization）不变。也就是说，在每一替换阶段，系统的输入—输出关系、行为模式与主观报告均保持一致。设想中，在仅替换一个神经元时，个体的意识状态不会发生变化（前提一）；如果每次替换都不引起意识变化，那么完整替换整个大脑也不会中断意识（前提二）。由此可得出结论：只要功能组织保持不变，意识状态亦将保持不变。换言之，意识随附于（supervene on）功能结构。查尔默斯称之为“组织不变性原则”。<sup>[3]</sup>

需要说明的是，这里所说的功能组织是指系统各组成部分之间因果互动的抽象模式。这种结构不仅涵盖系统内部各子部分的相互依赖，也包括系统与外部输入输出之间的因果关联。具体而言，功能组织由以下三个要素决定：（1）组件集合：系统被划分为若干抽象组件，每个组件代表一个功能单元；（2）状态空间：每个组件具有一组可能状态，可为离散或连续；

(3) 依赖结构:系统规定每个组件的当前状态如何取决于其他组件的先前状态及系统输入,并规定输出如何依赖于组件的历史状态。总之,功能组织刻画的是结构中的因果依赖模式。

有了上述交待,我们便可以分析神经元替换论证的合理性及其背后的理论承诺。前提一之所以看似合理,是因为:如果仅替换一个神经元,而替换元件在功能上完全等价,我们就没有理由认为这类微小变化会引发意识状态的剧烈转变。日常生活中我们不断经历神经元的自然死亡与重建,但意识状态并未因此中断或瓦解。因此,若承认大脑在生理更新中能维持意识连续性,就难以否认功能等价的人工神经元替换也能维持意识稳定。若拒绝这一前提,就必须接受一种极端立场:即使最小尺度的结构替换也可能立即摧毁或改变意识,这不仅缺乏经验支持,在理论上也显得极不自然。

前提二同样难以否认。否认前提二只有两种方式:(1)在神经元替换的某一个特定时刻,意识突然“跳跃”或“终止”;(2)在神经元替换过程中,意识体验不是瞬间中断,而是逐渐减弱。但这两种情形都不太可能。先看第一种情形。在系统行为完全不变的情况下(功能组织不变),意识的突然“跳跃”或“终止”既不可经验观察,也无法被理论定位。并且,这种情形承诺了一个节点,在这个点上意识突然“跳跃”或“终止”。然而,承认节点的存在会导致任意性。为什么是第 $n$ 个神经元的替换导致意识的消失,而不是第 $n+1$ 个?对于这个问题,任何回答似乎都是任意的。

第二情形意味着,一个原本拥有清晰视觉感受的主体,其经验可能逐渐变得模糊、黯淡,直到最终完全消失,但其语言报告与行为反应仍持续表达“我看见了红色”“我感受到疼痛”等主观状态。这一结论极具悖论性。若我们接受它,就必须承认一个理性、自洽、认知完备的系统在主观体验的判断上持续错误,即意识与认知彻底脱节。这不仅违背了我们关于意识与自我报告存在基本耦合关系的直觉,也意味着必须接受一种高度怀疑论的立场:日常神经变化(如神经元死亡)可能正悄然剥夺我们的

意识,而我们却全然不觉。这显然难以接受。

神经元替换思想实验之所以有如此强大的理论支撑力,不仅在于其直觉基础,更在于它形象演绎了计算功能主义的意识观。它表明,意识的存在与特性取决于系统中各部分之间的因果组织,而不依赖于这些部分的物质基质(substrate)。只要每个组件的当前状态与其他组件的先前状态及系统输入之间的决定关系,以及系统输出对组件历史状态的依赖方式保持不变,系统的意识特性就不会改变。这正是计算功能主义的基本观点:意识是一种功能性计算状态,其本质在于系统的功能组织,而非其物理基质。<sup>[12]</sup>

计算功能主义融合了两种传统思想资源:功能主义与计算主义。前者认为,心理状态是系统中具有特定因果角色的状态,其性质由在整体功能网络中的位置决定,而非其物理构成。后者将心灵视为信息处理系统,运行机制可抽象为图灵机或其他形式系统。图灵机的核心思想是,计算过程可完全由一组状态转换规则、有限的输入输出符号和明确的因果依赖结构定义。也就是说,计算的本质与物理实现无关:只要保持操作规则和符号系统,材料可以替换。计算功能主义将这一物理中立的抽象结构应用于心理领域,主张意识系统亦是一种“机制可迁移的图灵机”。在此观点下,心灵是一种高度复杂的计算系统,其状态变化服从一套形式化的函数规则,而意识则是该系统运行时实现的某种“高阶状态”或“过程轨迹”。

这一思想背后隐含着两个强有力的哲学信念。其一,计算本质上独立于物质基质:功能组织只需通过某种实现即可,不要求具体材料。就像同一套软件可由不同计算机执行一样,同一种心理状态,譬如疼痛,可为不同构造的系统所拥有,譬如人类、兽类、鱼类,甚至是外星人。其二,心灵状态是可同构迁移的:只要系统在组织层面与原系统同构,其意识状态也将一致。这两个信念的结合,使得上传体若保持与原脑功能同构,就被认为“有意识”。

然而,计算功能主义会导致一种有问题的二元论。这是因为,如果心理运作只是对形式

符号的计算操作，那么它便与大脑没有本质联系——大脑不过是众多能实现该程序的物理系统之一。这种二元论并非传统的笛卡尔实体二元论，即物质与心灵是两种独立实在，但仍具笛卡尔式特征：它坚持认为心灵的心理特性与大脑的物理属性之间没有内在联系。在这一框架下，心灵只是某种结构性的“高阶模式”，而非根植于物理过程的现象。这种将意识与实现条件解绑的立场，难以解释为何意识只在某些物理系统中生成，而在其他系统中则始终缺席。形式与物质的脱钩，正是当代意识理论面临的核心张力之一。

#### 四、可朽计算与意识的实现条件

计算功能主义所蕴含的形式与物质脱钩的思想，实际上源自经典的计算理论，尤其是图灵机模型及其对计算过程的定义。在图灵机的框架中，计算被设想为一组稳定、封闭、可重复实现的函数变换规则，计算过程被视为一系列符号的操作和状态转移，它们对物理实现具有中立性与可迁移性，只要满足形式结构与因果依赖关系的条件，任何物理系统——无论是碳基、硅基还是其他基质——都可以被视为该计算过程的实现者。计算由此被抽象为与时间、空间、能量和物质无关的形式结构，计算过程一旦被定义清楚，它便具有跨时间、跨平台、跨材料的生存能力。这样的计算，我们可以称之为不朽计算（immortal computation）。

不朽计算概念在认知科学与人工智能的早期历史中极具启发性，并被功能主义心灵理论采纳，用以说明心理状态与其实现者之间的可分离性，即功能主义所强调的多重可实现性。<sup>[13]</sup>然而，作为心灵理论的多重可实现论题却是一个形而上学上强烈但经验支持薄弱的主张。它在很大程度上依赖于直觉性设想，而非基于对心理实现机制的系统经验研究。在神经科学实践中，心理功能的分类与神经结构往往是对应的；也就是说，心理状态的界定并不是在脱离生物基质的前提下独立完成的，而是在对神经机制的持续研究中逐步形成和修正的。

最具代表性的例证之一是意识神经相关物（neural correlates of consciousness, NCC）研究的发展。NCC旨在识别在特定意识经验出现时系统性激活的神经结构或动态模式，其目标并非抽象界定“意识”，而是通过实证手段建立不同意识状态与其物理基质之间的精细对应关系。<sup>[14]</sup>例如，视觉意识被反复发现与背侧和腹侧视皮层中特定区域的活动密切相关，而某些前额叶区域的活动则更可能涉及注意或报告机制而非意识本身。这类研究促使神经科学家与哲学家对“视觉意识”与“视觉处理”进行更细致的区分，也推动了对意识范畴的再界定。此例充分说明，心理功能并非脱离物质、可任意实现的“形式结构”，而是深嵌于生理结构及其动力学中的具体现象。

既然心理功能的确定与神经结构及其动力学密不可分，那么我们就有理由怀疑：心理现象可能不仅仅依赖于抽象的功能组织，更根本地依赖于一类特定的物理计算过程本身。而这类计算过程，可用可朽计算来刻画。可朽计算指的是计算过程依赖于其物理基质的计算：当基质停止工作时，计算也随之终止。<sup>[15]</sup>

可朽计算反映的是一种深度嵌入其物理实现中的计算机制。这类计算不能被从具体的物质结构中抽象出来，也不能在平台之间无损迁移。与经典的不朽计算不同，可朽计算强调的是，实际物理系统进行高效计算的能力，往往依赖于对硬件本体特性的学习性适应，而这种适应性的成果无法被抽离为跨平台共享的形式规则。在这种模型中，连接权重、非线性反应函数以及输入-输出映射等关键参数，不是预先设定的，而是在具体硬件实例的限制条件下，通过与环境的持续交互、训练与反馈机制逐步调节而获得的。这些参数对外是不可迁移的，对内是不可抽象的；一旦其所依赖的物理结构损毁，整个计算过程也随之终止。由此构成的计算过程本质上是“可朽”的：它与特定的物理组织共生共灭，其运行依赖于特定材料的时变性质、局部扰动响应以及能量与信息流动过程的稳定维持。

可朽计算具有三个核心特征：耗散性

(dissipativity)、特异性(idiosyncrasy)、退化性(degradability)。<sup>[15]</sup>这三个特征与我们在经验层面观察到的意识现象存在显著的对应关系,因而有理由认为,可朽计算比传统的抽象计算模型更贴近意识的实际生成机制,能够为理解意识的实现条件提供更具解释力的理论框架。

耗散性意味着系统无法在热力学平衡状态下运行,而必须持续依赖能量输入与熵输出以维持其结构和动态行为。生物大脑并不是一个封闭的离散符号处理系统,而是一个处于非平衡态的耗散结构,其计算活动通过离子泵维持膜电位、通过代谢活动提供神经递质、通过脑血流调节维持局部能量供应。这些耗散过程不仅为神经计算提供能量支持,更与意识状态本身的维持密切相关。例如,在意识丧失状态(如深度麻醉或昏迷)中,大脑的能耗水平显著下降,皮层与丘脑之间的动力联结被破坏,说明意识并非与能量过程无关的形式计算,而是一种与代谢动态强相关的物理现象。<sup>[16]</sup>

特异性强调每一个系统——例如每一个大脑——在其物理构成、历史演化与环境互动中形成了不可复制的结构路径与参数配置。在可朽计算中,系统的有效参数配置(如连接权重与非线性响应函数)并非预设或可移植的模板,而是通过与特定硬件实例的交互性适应逐步训练而成。这一训练过程高度依赖个体物理结构的不可约性质,并最终将“计算能力”锁定在一个无法复制或迁移的物理基质之上。这一特征与意识的主观性在本体论层面展现出高度一致性。意识之为意识,正在于它总是某个具体主体的意识,具有不可还原的第一人称视角与经验结构。意识不仅是经验到某种东西的存在(如“有痛感”“有红色”),更是这种经验对于某个具体主体而言正在发生:它总是被某个主体所拥有,而不是仅仅发生于主体中。意识的这种“为我性”(for-me-ness)不是对系统功能或行为的外在描述,而是一种从内部生成的经验视角。<sup>[17]</sup>恰如可朽计算所依赖的结构与参数只能在特定物理实例中生成并生效,意识的主观视角也只能在某一具体、具身化的系统中浮现。

退化性指的是,系统的计算能力随着物理

结构的损耗和环境条件的恶化而不可逆地衰减,最终走向终止。与不朽计算的无限可持续、可重复调用不同,可朽计算始终嵌入在不可逆的物理时间进程中。它依赖能量输入、组织维持与物质更新,任何一项机制失效都可能导致系统功能的不可逆崩解。这种退化不是外在扰动导致的偶发终止,而是计算运行于真实物理世界中所必然面临的热力学命运。意识作为一种嵌入生理过程的现象,也体现出与可朽计算一致的退化性。我们不会在静止结构中发现意识,而只会在不断调节、持续耗散、逐步衰变的生命系统中观察到其暂时存在。在睡眠、昏迷、退行性疾病乃至死亡过程中,意识状态呈现出典型的边界模糊性与脆弱性。这种可被感知与追踪的退化过程,并非意识的偶发特征,而是其作为现象的常态结构之一。<sup>[18]</sup>

如果我们接受意识是一种可朽计算的观点,那么心灵上传的设想将陷入一个深刻的本体论悖论:一方面,上传体若要具备意识,就必须具备可朽计算所要求的三项核心特征——耗散性、特异性与退化性;但另一方面,正是这三项特征决定了上传体的意识只能是其自身的,而不可能与原体保持同一。因此,心灵上传若要实现“意识的延续”,恰恰会因为其对意识的真实实现,而失去了“同一性”的可能。

这种悖论并非是逻辑的冲突,而是意识的形而上学的内在张力的显现:它揭示出意识之所以为意识,正是因为它根植于一个朝向终结、无法复制、不可抽象的存在方式之中。意识的形而上学决定了,它的生成不仅要求复制某种功能组织,更要求维持一个具备耗散性、特异性与退化性的动态计算过程。这些特征无法被还原为结构等构,也无法在任意物理平台上稳定迁移。

无论是脑机接口技术、全脑仿真方法,还是人工意识,若采用的是不朽计算,即便在上传体中模拟了神经连接图谱,也会因缺乏与个体物理演化路径相对应的训练过程与物质耦合而失去意识。计算活动依赖对抗熵增的物理过程,一旦能量中断、结构破坏、调节机制失衡,意识所依托的计算结构便将不可逆地崩解。即

使上传系统能在短期内复制部分意识相关功能，其缺乏真实耗散机制与代谢支持系统的特性，仍可能导致系统稳定性不足，难以维持意识所需的热力学非平衡状态。而意识本身是在生理稳态中维持的暂时过程，一旦该稳态被打破，即便仿真仍在运行，意识状态也可能随之终止。因此，上传不仅面临结构复制的问题，更面临过程再现与维持条件重建的难题。

从可朽计算的视角来看，神经元替换论证隐藏着一个关键问题：它误将意识视为可在功能层面持续复制的稳定结构，而忽略了意识赖以生成的可朽计算之基质嵌入性。该论证设想，个体的神经元可一个接一个地替换为功能等价的人工元件。然而，如果意识的实现依赖于特定神经元之间在生理、代谢、时序上的耦合过程，那么替换便不可能是功能等价的。即便是渐进式替换，也会逐步剥离支撑意识生成的物理基础。某一神经元之所以能实现某一因果功能，并非仅因其处于图式结构中，而是因为其电生理特性、分子响应机制与周边支持结构（如微血管系统、胶质介导的稳态）恰好满足实现该功能的条件。一旦将其替换为物理机制迥异的硅元件，即使在短期内实现电信号响应匹配，也难以维持长期的稳定调节与退化响应。<sup>[18]</sup>因此，尽管功能等价的神经元替换在形而上学上是可能的，但在律则层面上却无法实现可朽计算的物性要求。

渐进式上传论证之所以表面上具有说服力，是因为它用替换过程的连续性来遮蔽意识实现条件的本体断裂。然而，一旦我们承认意识依赖于可朽计算的特异性——即它只能在一个具体系统内部，通过独特的物理结构、能量调节机制与历史耦合路径维持——那么就必须承认：哪怕替换过程再细致、再渐进，也始终在悄然瓦解原系统赖以维持意识的独特结构。一旦这种结构被完全替换，上传体的意识即便在表征上再相似，也只能属于一个新的意识系统，而不可能是原主体的延续。渐进式上传不是真正保留了“我”，而是以一种逐步重构的方式制造了另一个“它”。

也许有人会反驳说，对渐进式上传论证的

上述批评同样适用于大脑神经元的正常代谢过程，于是我们仍将面临前述极为反直觉的结论：我们可能因日常神经变化而不断“死亡”。然而，这种类比忽略了一个关键区别：在正常代谢过程中，意识所依赖的可朽计算机制并未被替代，而是通过系统内部的自我维持与自我修复动态保持下来。换言之，尽管神经元持续更新，大脑整体并未丧失维持意识所需的特异性调节结构；恰恰相反，正是这些微观变化在有机系统边界内被整合与调和，构成了意识主体连续存在的条件。

可朽计算的特异性并不排斥物理变化，而是强调意识的生成依赖于一个与自身退化过程持续交互的自调节体系。在自然生命过程中，这种调节是具身的、历史性的、嵌入性的，并不破坏系统的个体性基础。而在上传或替换的情境中，原始的自组织机制被外部工程性替代，个体性路径被中断。换言之，日常生理变化之所以不导致身份断裂，是因为它们发生在特异性得以维持的边界条件之内；而上传替换之所以构成断裂，是因为它打破了这一边界，使意识的实现条件从“被维持”变为“被重构”。即使替换是渐进的，其对系统可朽结构的破坏仍是累积且不可逆的。当重构积累到一定程度，便会引发有宏观效应的身份断裂。问题不在于系统是否发生变化，而在于这些变化是否越过了维持意识个体性所需的特异性边界。

正如我们所看到的，可朽计算揭示出意识实现的自体论约束：意识既不是游离于物质之外的结构性属性，也不是可复制粘贴的功能图式；它是一种受限于具体物理基质、历史路径与耗散条件的自体过程。因此，任何将意识从其生命性与死亡性中抽离的设想，不仅在技术上构造了一个不可证实的模型，更在自体论上误读了意识的生成逻辑。尽管心灵上传描绘出一幅战胜死亡的图景，但其逻辑前提正是对意识的非物质化误解——它试图用结构连续性取代具身特异性，用功能等构掩盖实现条件的不可迁移性。而一旦我们承认意识是一种可朽计算，承认其生成依赖于独特的物理、演化与退化路径，就必须正视上传的悖论：若上传无意

识, 则其延续毫无意义; 若上传有意识, 则其意识只能属于“它”, 而不再属于“我”。这不是技术悲观论, 而是对意识本体论地位的清醒把握: 意识的可朽性不是其局限, 而正是其存在的根基。

## 结 语

超人类主义为我们描绘了一幅美好的未来图景: 一个心灵摆脱肉身限制、意识在人工基质中延续的时代。而可朽计算揭示了意识生成的根本条件: 它不是游离于物质之外的功能结构, 而是依赖特定物理组织、自我调节与退化机制的暂时性现象。意识无法通过结构复制实现迁移, 它只能在具身系统的耗散维持中得以保持。若上传欲保留意识的真实性, 就必须承认其可朽性; 若试图剥除其可朽性, 留下的将只是一个失去主观视角与退化路径的空壳。

从可朽计算的视角看, 意识之所以宝贵, 正因为它不能被迁移、冻结或永生, 而只能被维持、经历并最终走向终结。面对心灵上传的技术幻象, 我们或许应当意识到, 意识的不可复制性并非一种缺陷, 而正是它彰显自己的方式。我们之所以是我们, 正因为意识只能在一次性的可朽结构中被拥有, 而不能被重现。因此, 与其幻想脱离终结, 不如反思我们如何在可朽之中成为独一无二的存在。这也许是面对心灵上传这一技术神话时, 哲学所能给予的最深刻的提醒。

## [参考文献]

- [1] Koene, R. 'Feasible Mind Uploading'[A], Blackford, R., Broderick, D. (Eds.) *Intelligence Unbound: Future of Uploaded and Machine Minds*[C], Hoboken: John Wiley & Sons, 2014, 90-101.
- [2] Mridha, M. F., Das, S. C., Kabir, M. M, et al. 'Brain-Computer Interface: Advancement and Challenges'[J]. *Sensors*, 2021, 21(17): 5746.
- [3] Chalmers, D. 'Absent Qualia, Fading Qualia, Dancing Qualia'[A], Metzinger, T. (Ed.) *Conscious Experience*[C], Paderborn: Schoeningh-Verlag, 1995, 309-328.
- [4] Chalmers, D. 'The Singularity: A Philosophical Analysis'[J]. *Journal of Consciousness Studies*, 2010, 17(9-10): 7-65.
- [5] Kurzweil, R. *How to Create a Mind: The Secret of Human Thought Revealed*[M]. London: Penguin, 2013.
- [6] Schneider, S. *Artificial You: AI and the Future of Your Mind*[M]. Princeton: Princeton University Press, 2019.
- [7] Chalmers, D. 'Uploading: A Philosophical Analysis'[A], Blackford, R., Broderick, D. (Eds.) *Intelligence Unbound: Future of Uploaded and Machine Minds*[C], Hoboken: John Wiley & Sons, 2014, 102-118.
- [8] Piccinini, G. 'The Myth of Mind Uploading'[A], Clowes, R., Gärtner, K., Hipólito, I. (Eds.) *The Mind-Technology Problem: Investigating Minds, Selves and 21st Century Artefacts*[C], Cham: Springer, 2021, 125-144.
- [9] Goldwater, J. 'Uploads, Faxes, and You: Can Personal Identity Be Transmitted'[J]. *American Philosophical Quarterly*, 2021, 58(3): 233-250.
- [10] Olson, E. *The Human Animal: Personal Identity Without Psychology*[M]. Oxford: Oxford University Press, 1997.
- [11] Cerullo, M. 'Uploading and Branching Identity'[J]. *Minds and Machines*, 2015, 25(1): 17-36.
- [12] Polger, T. 'Computational Functionalism'[A], Robins, S. (Ed.) *The Routledge Companion to Philosophy of Psychology*[C], London: Routledge, 2019, 148-163.
- [13] Funkhouser, E. 'Multiple Realizability'[J]. *Philosophy Compass*, 2007, 2(2): 303-315.
- [14] Aru, J., Bachmann, T., Singer, W., et al. 'Distilling the Neural Correlates of Consciousness'[J]. *Neuroscience & Biobehavioral Reviews*, 2012, 36(2): 737-746.
- [15] Ororbias, A., Friston, K. 'Mortal Computation: A Foundation for Biomimetic Intelligence'[J]. arXiv preprint, arXiv: 2311.09589, 2023.
- [16] Friston, K., Buzsáki, G. 'The Functional Anatomy of Time: What and When in The Brain'[J]. *Trends in Cognitive Sciences*, 2016, 20(7): 500-511.
- [17] Zahavi, D., Kriegel, U. 'For-me-ness: What It is and What It is Not'[A], Dahlstrom, D. O., Elpidorou, A., Hopp, W. (Eds.) *Philosophy of Mind and Phenomenology*[C], London: Routledge, 2015, 36-54.
- [18] Cao, R. 'Multiple Realizability and the Spirit of Functionalism'[J]. *Synthese*, 2022, 200(6): 506.