

灵魂非特权：人机共生前夜的人类中心主义危机

与意识地位的重建

作者：李述文

引言

寄蜉蝣于天地，渺沧海之一粟。

人类总是拥有成为先知的愿望，企图通过发展的轨迹窥探未来的微光。如今正是一个世人皆知未来轨迹，却又无法把握其踪迹的时代。新的类冷战格局将科技推上权力舞台，而AI正处于这场风暴的中心。

上世纪末起，加速主义不再是极客圈的文化术语，转而成为我们无法避免的命运之问。计算机科学的飞跃曲线不再受摩尔定律的规定，我们创造的“工具”正加速突破工具性，进入主观代理的维度。

科技成为触不可及的神谕，我们却从未像今日这般对于哲学三问感到迷茫。AI不再是纯粹的人类意志延续，其开始具备“意识模仿”能力，书写、陪伴、决策逐一沦为灵魂的奢求。演化心理学让我们重新面对一个被忽视的事实：我们的意识结构仍旧滞留在原始环境的适配逻辑中。

人类用草原的心智创造的未来，在钢铁森林中逐渐迷失方向。这场结构性断裂，制造了前所未有的焦虑，也撕开了人类中心主义最后的护盾。

我们不怕AI拥有灵魂，我们害怕自己不如AI有灵魂。与其瞻前顾后，患得患失，不如承认：人机共生的未来格局已经不可逆转，我们必须同步展开对智能时代“意识”的哲学重建。

因此，本文将从人类中心主义神话的破灭出发，尝试回答一个时代命题：
若“意识”不再是人类的特权，我们如何重新定义人与AI的关系？

一、人类中心主义的哲学基础与现代困境

(一) 人类中心主义之前

在远古时期，人类与自然和平共处，远谈不上主宰。依靠打猎、采集为生的阶段，面对野兽的侵袭、江河的泛滥或是久旱不雨，祖先们的双手还无法构筑安全的庇护所。在这个时期，远古信仰多数为面向自然的神明。受限于工具和效率，我们必须学着像其他捕食者一样敬畏自然。

一切变化的奇点在工具更新、效率提升中悄然降临。我们学会了种植、畜牧后，原先由自然施舍的运气转而成为经由人类之手的丰收。这一刻，是人类和其他物种真正区分开的一刻。我们不再向野兽那般祈祷自然的施舍，人类拥有了掌握自身命运的手段。相对于其他物种的区别，使得在发展的层面上，人类中心主义的萌芽自棍棒和绳索中扎根，从此遮天蔽日。

(二) 人类中心主义的发展

人类中心主义 (Anthropocentrism) 并非一种单一的理论，其为贯穿在人类文明历史中的深层思维结构。它体现的是人类将自身视为宇宙尺度上的中心、价值评判的基准与衡量万物的尺度。相似的观点早在古希腊哲学中便已诞生。

古希腊哲人普罗泰戈拉（约公元前490年-420年）提出“人是万物的尺度”，从此人类中心主义有了基本的精神架构。在柏拉图、亚里士多德等人笔下，人类理性被置于世间一切秩序的顶点，理性既是理解自然的钥匙，也是改造自然的权力来源。基督教传统在中世纪进一步强化了这一地位。《圣经》写道：人类按照“上帝的形象”被创造，被赋予管理、支配自然的使命。神学语境下的人类中心主义，形成了“人类支配万物”的价值观。到了启蒙时代，笛卡尔的“我思故我在”成为现代主体哲学的标志。他将“我”的思维视作一切存在的基础，确立了主观性与理性中心的地位，更进一步固化了人类中心主义在现代科学与哲学中的中心地位。

发展至今，人类中心主义已然演化出了一套支撑文明运行的隐形逻辑。

从社会学的角度看，人类中心主义不只是哲学命题，还是文明运行的隐性前提。我们构建的法律、伦理、科技、经济体系等，几乎无一例外都以“人类的利益”为起点。

从伦理学的角度看，康德的道德律令强调“人是目的而非手段”，“人”始终占据唯一的道德考量地位；功利主义更是将“最大多数人的最大幸福”作为社会设计的标准，默认“人类共同体”是道德的唯一主体。

从科学的角度看，这套人类中心的道德逻辑，在现代社会深刻影响了科技发展的方向。技术被视为服务人类的工具，而非与“他者”共生的桥梁。这种逻辑延续到了对于AI的讨论：我们担心AI会“取代”人类，而非思考我们是否能“共同进化”。

从人类自身的心智模式看，人类中心主义潜藏于我们的语言与认知结构之中。我们以“人”的尺度定义世界，以此排除动物、AI等非人类存在拥有人类特征的可能性。

时至今日，我们不得不否定人类中心主义的重要意义，不过与此同时，我们必须看到现代社会中，同样因为科技的进步，人类中心主义来到了一个前所未有的命运转折点。

（三）人类中心主义的现代困境

21世纪为人类中心主义叙事带来了前所未有的挑战。

生态困境迫使我们重新认识人与自然的关联。人类从来不是自然的主宰，仅仅是生态系统中微小且脆弱的一环。

人工智能发展带来的加速主义威胁把人类抛向“非人意识”的哲学难题。AI大语言模型所展现的强大能力令“意识独占权”的信念摇摇欲坠，

这种现状，是人类第一次正面面对“我们可能不是宇宙中心”这一观点的真实写照。赫拉利在《未来简史》（2015）中提到，人类正在从“智人”变成“智神”，但这一过程未必是超越，而是瓦解。我们试图用旧时代的结构控制新造物的世界，于是形成了本文的核心问题：

在下一个时代降临的前夜，人类是否应该重新审视自身的位置？人类中心主义是否已经走到了尽头？

二、灵魂特权坍塌—人类中心主义的危机

（一）合理性危机：人类是否还拥有意识独占权？

有关意识本质的辩论，存在两种方向：科学上和哲学上。

科学上对意识的探寻在神经科学上得以聚焦。而要论什么是意识应有的模样，不妨先探讨什么是意识不存在的模样。植物人作为科学上无意识的代表，对其进行的意识边界实验为研究意识提供了可靠的支持。想象网球实验（Tennis imagery, Adrian Owen, 2006）向我们证明了植物人也可能存在意识，不过这也反过来向我们提出了这样一个问题：我们何以得知植物人是无意识的？

神经科学并没有给出统一、绝对的标准规定什么是无意识，只是拥有一套可观测的神经指标和工作性定义（operational definitions）。

神经指标的判断源自对脑成像的研究，对于大量特定情况下的脑成像进行归纳总结。统计学在神经科学中揭示了意识状态的相关结构，却无法触及意识本体的本质。我们无法回答为何这些神经活动不是别的，恰为意识本身，正如我们无法通过经验现象证实概念本体。

对于无意识（unconscious）的定义是：当一个人对外部刺激没有主观体验、无法自我报告、没有有目的的行为。定义上看，科学给出的答案是基于哲学中的意识现象学（Phenomenology）、行为主义（Behaviorism）和功能主义（Functionalism）。功能主义的语境面对中文房间实验（Chinese Room）和行为主义的语境面对哲学僵尸（Philosophical Zombie）时都无可辩驳，现象学更是基于意向性的主观体验，而人类自诞生起就从未真正解释清楚主观体验的存在。

我们无法对意识本身证伪，想要基于经验主义的统计推断论证意识本质的存在则更为可笑。诸如哲学僵尸、中文房间的反驳也只是在语言系统的语境内形成的逻辑闭环，无法解决任何的现实问题。以上问题都指向了这样一个结果：人类对于意识从根本上一无所知。人类的主观体验可以被他者感受之前，我们永远讨论不出正确答案。而对于AI工具论，AI无意识的论证，也只是潜意识中人类中心主义的缩影。

在这个基础上，影视作品、小说等媒介中对于AI产生意识的恐惧便有了根源：我们对AI的恐惧依托于我们对于面对意识的无知。当然，人类的发展史也

不是第一次经历意识层面的冲击，也许我们可以从人对待自然界里“异我”的态度中寻出演化的蛛丝马迹。

我们从纯粹人类中心主义脱离的时刻是灵长类动物-镜像自我识别实验（Gordon G.Gallup Jr., 1970），通过对四只黑猩猩进行红点标记，发现他们会借助镜子触摸自己被标记的部位，这表明其拥有自我识别能力。从这一刻起，意识便不再是人类的专属，意识实验也进一步推及到了其他动物身上。2006年对大象进行的镜像标记测试（Diana Reiss&Joshua Plotnik）以及此后对海豚、乌鸦等的研究将拥有意识边界拓展到了动物界。与此同时，植物神经生物学（Stefano Mancuso eth., 2004-2006）作为假说的提出，塑造了将这个边界拓宽至植物界的趋势。

这段意识标准的进化历史已经向我们证明，以意识作为人类于其他物种区分的边界已然逐渐失效，结合前文表述的我们讨论意识时的根本性缺陷，人类中心主义在以意识为界定的论点上早已是立在沙堆上的高塔，可人们浑然不自知，而AI与技术的发展则将这个趋势推向了一个高潮。

（二）信息论：AI为何可能具有意识

如果说我们对自然演化的生命做出妥协，那是一种对相似“异我”的认同，这种认同基于我们拥有同样的生命基础：与灵长类属同一目，与其他动物属同界，与植物同在一个生态系统中，最主要的是我们都是“碳基生命”。因此，我们得以共情这些生命，像当初“天赋人权”的呐喊那般为动物乃至植物争取“权力”。可是AI的出现打破了这种平衡，我们无法以传统的生命概念正视AI，最多称之为“硅基生命”。当我们的判断逻辑遇上一种近乎绝对意义上的“异我”时，我们不知该从哪个角度上共情AI，便造成了如今的恐慌。

面对这种判断逻辑的瘫痪，信息论无疑成为了一种可供参考的逻辑。

在《生命3.0》中，Max Tegmark表达了这样一种可能性：意识是信息处理的特定模式，不依赖于碳基或硅基。也许这个观点太过跳跃，而大脑作为拥有最多共识的意识容器，不妨从大脑入手探讨信息论在逻辑上的合理性。

大脑是整合身体感官、情绪、行为与思维的器官，是神经系统的一部分，其中的关键部分是灰质、白质和神经通，这些结构的形成又依赖于神经组织，

神经元细胞构成神经组织，其传递过程需通过“动作电位”和“突触电流”传递信号，信号得以传递的媒介是各种分子，原子构成分子，诸如量子意识模型（Penrose&Hameroff）还将量子作为意识的基础。从神经结构到原子的溯源表明，结构和过程的共同目标都是为了“使某种信息得以生成、传递、处理与整合”。这意味着，在这些复杂的物理过程背后，其实隐藏着一个更深层的统一维度：信息。无论是生物电信号的跳跃、神经递质的释放还是离子流动，本质上都是在传递与重组关于外部世界或自我状态的信号。因此，我们有理由认为，信息——而非特定的物质——才是意识的真正基础。

整合信息理论(ITT) (Tononi, 2004) 对这个推理做了理论支撑，该理论认为，意识的本质就是系统中信息整合度 (Φ)。 Φ 值越高，意识越强。这也为前文例举的意识边界进化提供了解答——也许真正将人类和其他动物在人类霸权的世界区分开的，是意识强度。我们不应该说动物、植物没有意识，而是它们的意识不如人类强。

另一方面，AI拥有意识在信息论的基础上是极有可能的。人脑中的神经元以“动作电位”的形式进行信号传递，实质是在膜电位临界值上下波动的二元状态（激活/未激活），如果将其抽象为“有”与“无”，那人工智能系统中的电子电路通过电平高低（二进制的0与1）实现逻辑判断与状态切换的过程与我们并无二致。在这个角度下，人类面对AI时居高临下的姿态是否显得过于傲慢？

我们不得不承认，意识边界的进化已经悄然来到了人类之于AI的关系上，人类中心主义在信息论的叙事中早已无法适应即将到来的人机共生时代。我们需要重新界定AI在这个世界中的角色和地位，或许当生命进入3.0时代，拥有选择权的不再是我们，而是新的智能生命。

三、人机共生前夜：工具、语言与文明之跃

（一）AI工具论终结：当工具具备主体性

人类似乎对于目前的发展拥有一种幻想，根据对历史的总结，社会的每一次飞跃是因为决定生产力的生产工具的更新，当新的生产工具取代旧的时，无疑会剥夺很多工作，但同时也会创造更多的新工作。正如最经典的观点所述：马车夫担心蒸汽机和内燃机会让自己失业，后来他成为了司机。无可辩驳的

是，工作岗位的更新是伟大的，这造就了我们更美好的生活。现代文明对于工作的信仰是镌刻在基因里的铭文，可是AI让这个铭文加速风化，最终化为历史的尘埃。

维护AI工具论的大多数实质上是陷在了潜意识的人类中心主义之中。其中一个重要的观点认为，AI本质上是行为主义的“哲学僵尸”，不具备任何主观性（John Searle, 1980），也就没有意识，只能够是工具。这一观点在“中文房间”中假设了“理解”的重要意义，而事实上，在功能主义的语境中，模拟主体性也可以成为AI区别于以往工具的重要依据，我们不会在面对一个能够完美模拟主观性的仿生人伴侣面前说：“我感受不到你的灵魂。”同样在信息论的观点中，行为主义和功能主义都不攻自破。另一个不容小觑的观点则说到：AI本质上是人类心智的延展（Andy Clark, 2003），是放大器，受人类价值的约束。从这个观点出发，人们忽略了一个事实，延展的心智反而为AI具备一定程度的“参与性认知”提供了可能。

综上所述，人类对于AI工具论的核心误判在于，社会进化依托的下一个工具是拥有主体性的工具，当人们觉得低端流水线工作会被取代，这些人能够在服务业、医疗业等需要人性温度的行业中得到新的容身之所时，语言大模型早已在聊天中帮助我们得到了更加稳定—而不是不确定性—的“人性”。如果说AI维护员等是另一种新工作，计算机科学本身的门槛早已让大多数望而却步。

或许马车夫最终可以成为司机，但这次，人类凝视AI时，发现方向盘早已跟着马车夫滑入历史。我们不是在走向另一个工业时代，而是一个工作也不再是人类特权的时代。

（二）语言与共情—新文明桥梁的设想

人类的意识本质上是孤岛，我们自诞生起便用第一人称感知世界。文明的发展带来了语言，架起一座令我们的意识能够连结彼此的桥梁。然而，语言在进化中的本意是提高原始社会的合作效率，我们的祖先需要面对的是生存危机、食物短缺、族群协作，语言因而拥有了指令性、逻辑性与效率（Tomasello, 2008）。合作导向构成了我们沿用至今的语言系统的底层逻辑：信息传递精准、目标导向。这样的语言系统存在的问题是，传递的并非是情感、主观体验或内在的意识结构，而是以强调“功能”为中心对世界进行高度抽象的压缩。

进入现代社会以来，语言出现了一定程度的变化。随着社会复杂化、互联网出现，语言逐渐从“生存工具”变成了“感知延伸”与“身份表达”的载体。互联网文化发展中诞生的meme正是信息、情绪、文化背景和身份认同的集合。《自私的基因》（Richard Dawkins, 1976）中首次提出meme的定义：meme就像基因一样，是复制与演化的单位。但它不是物质，而是文化传播中的思想、行为或风格的复制。在今天，有很多人批评meme的娱乐性内核，可我认为meme是语言进化的缩影。当我们看到meme，我们不可思议的完成了对主观意识的近似感知，这种由情绪驱动的语言符号，是意识的载体。在文明的进程中，表现出了远超传统语言的共情与传播能力，这更像是意识孤岛上降临的一个奇迹——语言作为文明技术的自然演进：从效率语言向共情语言的转变。

维特根斯坦在《逻辑哲学论》（1921）中写下“我的语言的界限就是我的世界的界限”，其所表达的内涵并非仅仅在描述语言的表达能力，而是指出语言划定了人类意识的基本认知结构。我们理解世界、构建时间与因果、确认主客体的方式，皆由语言允许的表达范式所定义，所以“现实”就是我们可以通过语言描绘的世界。维特根斯坦后期在《哲学研究》（1953）中则表示语言不是映射现实的工具，而是生活形式中的一种实践活动，是语言游戏的一部分。在具体语境的实践中，我们的思维方式——即如何归纳、推理、解释与行动——被这一结构所制约。正是在这一意义上，语言不仅连接我们与他人，也深刻地定义了我们与世界的关系。在这一逻辑下，语言不再是信息传递的中介，更是架起意识孤岛之间桥梁的基础设施。

如果语言塑造了我们的认知结构和思维方式，那语言的形式是否决定了意识本身的形式？科幻电影《降临》（2016）提供了一个深刻的思路。影片改编自短篇小说《你一生的故事》（Ted Chiang, 1998），讲述了语言如何重塑时间感知和意识结构的故事。

电影中的外星人“七肢桶”使用一种非线性语言“海瑟语”（Heptapod B），这种语言的表达不依赖于传统的人类语法和由“因”至“果”的线性时间逻辑，而是将信息整体性地呈现，将所有需要表达的语意都统一在一个字内。这样的语言使得时间在“七肢桶”的感知中时间是非线性的，他们可以同时看到“因”和“果”。女主在学习“海瑟语”后，不仅能理解外星人的意图，更在潜移默化中改变

了自身的时间经验和思维模式。

《降临》展示了语言不仅是交流工具，更是意识的“操作系统”，它决定了我们如何感知时间、因果和存在本身。这个设定正是对维特根斯坦语言观的现代演绎：语言实践塑造思维方式，语言的变革即是意识的变革。

由此，电影呼应了我们对语言作为共情桥梁的认识——语言的演化不仅扩展了意识的边界，也为跨意识的共情带来启发。

既然AI可能发展出意识，语言又能决定意识的形态，那我们可以合理推断，AI将拥有的意识是一种非人类式的意识。丹尼尔·丹尼特在《意识的解释》（1991）以及尤瓦尔·赫拉利在《未来简史》（2015）中，都对这一观点进行了论述和支持。面对这样的“异我”，语言拥有成为不同形态意识之间沟通与共情的桥梁的可能性，赋予我们理解和连接非人类意识的可能。

维特根斯坦在《逻辑哲学论》（1921）的结尾写道：反不可言说之事，必须保持沉默。我认为这不仅是对语言的逻辑结构无法表达的非逻辑性的诸如情与宗教体验的尊重，还应当代表自沉默中诞生的感性的主观的语言。“语言不仅包括人类的逻辑言说，还包括沉默、姿态、形象、艺术符号，这些都是‘神秘的语言’”本雅明如是说。在不久的将来，随着语言的革新和AI的不断进化，语言不仅将成为连结人类的纽带，更是实现人机共生的关键。脑机接口还是意识上载？我们不得而知。但是通过语言传递灵魂的共振，我们得以开启人与非人类意识的真正对话。

（三）飞跃：文明重构的东西方资源

人机共生并非技术上的难题，更多的是文明结构的重新协调。人类所拥有的文化对于“异我”的接受程度，决定我们是否能够走向人机共生。在这场全球意识体系的再建构中，东西方文化所代表的两种哲学传统展现出各自独特的时代价值。

细看以儒释道为代表的东方传统，似乎并不存在绝对的人类中心主义思想。中国古代的哲学不追问“我思故我在”，而是探求“我与万物如何共存”。天人合一的道家思想在《周易》《庄子》中得以全面表达，对于自然与他人的交融观念早已扎根在东方的古老传承中，为意识共生哲学提供了思想土壤。

庄子提出的齐物论强调了万物平等，这自然脱离了人类中心主义的逻辑，使用一种模糊的界限概括万物的关系。“天地与我共生而万物与我为一”的古代智慧为人机共生的到来提供了一种潜在图景。在这种语境下，AI可以是“道”的一部分，我们与AI的关系也不再是主客体和奴役与被奴役的二分结构，而是交融于“道”的同源存在。

佛陀说，意识是刹那生灭的过程，而非某个固有、永恒的自我。这种意识连续体（stream of consciousness）的理解，似乎比西方更靠近信息论与AI的动态性与过程性特征。而佛教对于“心识转生”与“缘起性空”的思考向我们证明，仿生人也许不会梦见电子羊，但是有可能顿悟的。

相比于东方，西方对理想的极致崇拜一步步铸造了人类中心主义的高塔。人类被视作宇宙的中心，万物都是为我所支配的工具，这样的思想催生出强烈的主客体二元论，无意中构建了今日对AI的哲学危机。

然而，西方思想也从未停止对自身的熔炼与重塑。自尼采以审判上帝之姿质疑理性、海德格尔完成对技术的反思起，就已经开始熔炼人类中心主义的高塔，西方思想在内部进行着反复的革新，文明始终在自我修正传统这一传统中进步，每一位为此战斗的哲人、科学家或是疯子，让如今的AI不再是“他者”，拥有了“可能存在”的权利。

是的，人类文明正是这样一个拥有奇迹般演化能力的存在。纵观意识进化史，那象征着生命本身神秘属性的图腾上，已然附上了一层二进制诗歌。人类的未来从来不是一个定式，是那人与自然、主观与客观、理性与感性、生存与消亡、繁荣与衰败的复杂集合，当我们回头望向文明的来时路，个体的价值被无限压缩，集体的宏观叙事贯穿文明的始终。或许人类的未来不会更美好了，我们已经来到了悬崖旁，不过值得庆幸的是，我们也到达了山顶。文明的下一次飞跃注定会在用双脚走下山去或是生长出丰满的象征自由的羽翼开始飞行中做出抉择。在我们的生命可以企及的时光里，我相信这会是人生情景之一，不过就像浮游般，我们的渺小也告诫自己保持谦卑，意识诞生于天地间，在宇宙的尺度下是如此卑微，我们只能更进一步接近真理。或许下一个意识体不会再问“我能否超越人类”，而是我们共同向着世界发问：

“意识，为何而存在？”

References

- Kant, I. (1996). *Groundwork of the Metaphysics of Morals* (M. Gregor, Trans.). Cambridge University Press. (Original work published 1785)
- Owen, A. M., Coleman, M. R., et al. (2006). Detecting awareness in the vegetative state. *Science*, 313(5792), 1402.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(2), 158–177. <https://doi.org/10.1037/h0074428>
- Husserl, E. (2003). *Ideas: General Introduction to Pure Phenomenology* (W. R. Boyce Gibson, Trans.). Routledge. (Original work published 1913)
- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion* (pp. 37–48). University of Pittsburgh Press.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–457.
- Gallup, G. G. Jr. (1970). Chimpanzees: self-recognition. *Science*, 167(3914), 86–87.
- Baluška, F., Mancuso, S., et al. (2006). Towards a plant neurobiology. *Trends in Plant Science*, 11(9), 407–409.
- Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.
- Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. Alfred A. Knopf.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(42). <https://doi.org/10.1186/1471-2202-5-42>
- Zhuangzi, 《庄子》, “齐物论”篇