



# 关于AI原生的几点探讨

## Several Insights into AI Nativism

何宝宏/HE Baohong

(中国信息通信研究院, 中国北京 100191)  
(China Academy of Information and Communications Technology, Beijing 100191, China)

DOI: 10.12142/ZTETJ.202506007

网络出版地址: <https://link.cnki.net/urlid/34.1228.TN.20251219.1010.004>

网络出版日期: 2025-12-19

收稿日期: 2025-10-15

**摘要:** 人工智能的应用模式主要分为两类: 一是助力传统产业的升级改造, 实现存量系统的智能化; 二是依托智能原生催生出新技术、新产品和新服务, 构建全新的智能系统。智能原生的概念虽已被广泛接受, 但相关探索仍处于相对初级阶段, 其理论内涵还不够丰富。本文从技术、业务和人三个角度, 探讨对智能原生的观察和思考, 并展望其发展趋势, 希望能对AI的应用落地和模式创新起到一定的参考作用。

**关键词:** 智能原生; 原生多模态; 数据战略; 智能算力; AI人

**Abstract:** The application modes of artificial intelligence are mainly divided into two categories: one is to facilitate the upgrading and transformation of traditional industries and realize the intelligent transformation of legacy systems; the other is to spawn new technologies, products and services based on intelligent nativism, so as to build brand-new intelligent systems. Although the concept of intelligent nativism has been widely accepted, relevant research is still in a relatively preliminary stage and its theoretical connotation remains insufficiently enriched. From the three perspectives of technology, business and human beings, this paper discusses the observations and reflections on intelligent nativism, looks forward to its development trends, and hopes to provide a certain reference for the application implementation and mode innovation of AI.

**Keywords:** intelligent nativism; native multimodality; data strategy; intelligent computing power; AI human

**引用格式:** 何宝宏. 关于AI原生的几点探讨 [J]. 中兴通讯技术, 2025, 31(6): 49–52. DOI: 10.12142/ZTETJ.202506007

**Citation:** HE B H. Several insights into AI nativism [J]. ZTE technology journal, 2025, 31(6): 49–52. DOI: 10.12142/ZTETJ.202506007

**数**字化转型旨在通过互联网、云计算、大数据和物联网等信息技术的应用, 对企业等主体的业务流程、商业模式和运营管理进行全面重构, 以提高效率、降低成本和提升竞争力。这一全球性的浪潮已经持续了20余年, 未来几年仍将保持高速增长的态势<sup>[1]</sup>。

数字化转型是数字技术对已有系统的赋能, 而数字原生是相对于数字化转型而言的, 它更强调土生土长于数字时代并以数字技术为基础或核心的全新系统。数字原生(代)最初指那些从小就生活在数字化环境中, 天然地对计算机、智能手机和互联网等数字产品有亲近感的一代人。但经过20余年的发展演进, 数字原生的概念已经从专指一代人扩展到技术、业务和生态等数字世界的方方面面, 泛指那些从诞生之初, 就基于数字技术或以数字技术为核心的新业态<sup>[2]</sup>。

近年来, 随着以大模型为代表的生成式人工智能的发展成熟, 传统的数字化转型和数字原生的概念, 已难以适配人工智能(AI)时代的新变化与新需求, 亟需融入智能化和智

能原生的新动能。智能化是应用AI技术对存量系统的智能改造, 这一概念相对容易理解。与之不同, “智能原生”的概念虽然已经被普遍接受<sup>[3]</sup>, 但其关注的是AI带来的智能增量, 且理念更为超前, 相关研究还处于初步阶段。对智能原生的探讨可以从技术生态、科技伦理、数字文明等不同维度展开, 也可以选取更具体的切入点。本文选择从技术、业务和人3个方面, 探讨并展望对智能原生的一些理解。

### 1 AI原生技术

AI原生技术指那些本身为AI而设计, 基于AI第一性原理而设计、开发并运行的技术体系。本文将聚焦AI基础设施(AI Infra)层技术, 仅探讨原生多模态、原生数据和原生算力, 以及与智算中心配套的供配电和制冷技术。

#### 1.1 原生多模态

目前主流的多模态模型是多个单模态模型的简单拼凑。

这类模型通过独立训练文本、图像、语音等子模型，再在后期实现子模型间的协同对接以完成多模态交互。这种拼接式的多模态架构，使得模态间信息传递依赖于手工设计的特征工程，容易导致语义关联丢失，也限制了模型的跨模态推理能力，使其难以应对复杂的应用场景。

原生多模态模型首先需要进行架构层面的创新，以深度融合各种模态。比如，不同模态需要共享参数空间，构建统一的表征空间，将不同模态数据映射至同一语义维度；引入跨模态的注意力机制与模态感知的权重分配机制，使模型能够动态聚焦关键信息；采用多模态联合优化训练策略，而非独立训练子模型，具体可以采用分阶段训练策略：先固定语言模型参数，再优化视觉编码，最后完成全模型的微调。

原生多模态除了需要实现以上架构层面的创新外，还需要在数据层面构建多模态数据库、在算力层面研发多模态芯片，并在跨模态信息的精准对齐与交互、多模态工具链研发、人机交互与用户体验等方面，开展更多的原生性创新实践。

## 1.2 原生数据

2025年，大模型训练“撞上”了数据墙已经成为业界共识<sup>[4]</sup>，即模型训练所需的高质量数据即将耗尽或难以获取，导致模型性能提升陷入停滞甚至出现倒退。互联网上可用于高质量训练的数据已接近开采殆尽，并且现有数据大多存在多样性偏差、内容过时或信息虚构等问题。与此同时，企业私域数据流通依然受阻，行业领域数据的版权壁垒也居高不下。

无论是公网数据、企业私有数据，还是行业专业数据，采集、清洗和保存这些数据的初衷，可能是为了信息传播、社交互动、风险控制或溯源审计等，但几乎都不是为了训练大模型。只是近年来，随着深度学习和大模型技术的发展，这些数据经过结构化重构、隐私增强和人工标注等处理后，被用于大模型训练，实现了二次开发和新价值释放。训练AI时缺乏“专业教材”的现象，使得AI训练领域正面临着转变数据原始意图时的供需错位、训练资料缺陷和成本高昂等挑战。

更深一层看，大模型所谓的深度学习，本质上并不是从物理世界直接获得或通过互动习得的一手知识，而是来自人类中介的二手数据，这类似于从“纸面知识”而不是从实践中习得知识的模式。AI的这种间接学习模式，会将其智能水平局限于人类认知的边界内。结合前面的讨论可以发现，目前AI训练实际上是基于人类提供的二手数据开展二次开发与利用。

为此，从短期来看，除了继续优化传统的数据“转型”之路外，还需要主动设计并开发面向AI的数据战略，即把模型训练作为数据生产的第一目标，重构数据全生命周期，形成数据、模型和应用的闭环生态。比如，在数据采集阶段，根据模型任务（如对话、图像生成）场景化定制数据采集方案，同步采集文本、图像、传感器等多模态数据，以避免后期模态拼接引发的语义断层问题；在数据清洗阶段，利用模型自身能力自动筛选数据和反向校正偏差；在数据保存阶段，像代码库一样管理数据集的版本迭代，在进行数据存储时就完成隐私增强保护，以避免后期再开展脱敏处理等冗余操作。

哈佛大学的一项最新研究发现<sup>[5]</sup>，在大模型训练数据中加入约10%的毒性数据，可以使毒性特征更集中且更易分离，更有利模型的长期稳健性。这为业界解决“高质量”数据问题提供了新思路：与其付出极高代价对数据进行彻底清洗，可能还不如引入生物学上的毒性免疫原理。AI大模型的输出是统计意义上的最优解，即使引入极少量的噪声数据（比如错误数据、毒性数据等）对输出结果的影响也很有限。

为此，从长远来看，模型训练应该实现“去人类中介”，通过构建多模态直接感知体系，更直接地触及世界运行的本质规律。比如，通过高精度脑机接口捕捉神经元集群的原始电信号，利用分布式触觉传感器网络，直接感知物体表面的分子排列与力学特性，结合环境交互反馈形成闭环学习，让AI在物理世界中自主验证认知假设等。

## 1.3 原生算力

在AI发展的早期阶段，相关运算主要依赖通用中央处理器（CPU）进行简单计算，AI与芯片彼此之间几乎没有相互影响。后来随着深度学习的兴起，AI模型的参数规模开始激增（2012年，AlexNet模型的参数规模已达6000万），通用CPU算力不足的问题开始凸显。而原来主要用于图形处理的图形处理器（GPU）凭借并行计算（包括软件部分的计算架构）的优势，被引入AI大模型的训练。但在引入初期，GPU仅被当作加速工具，经常未针对AI特性做深度优化。2017年之后，随着Transformer架构的提出、大规模模型训练的普及以及缩放定律（Scaling Law）的发现，AI模型的复杂度和算力需求呈指数级攀升。各类芯片厂商针对AI特性开始深度优化硬件架构，让原本主要服务于图形处理的GPU，快速演变成了以AI为核心用途的算力模式。

AI与算力的关系，已从早期的通用支撑、中期的辅助性加速器，发展到现在的原生芯片创新与AI算法优化循环

的共生关系。AI原生芯片针对AI算法的特性做了大量的定制化开发和深度优化。比如，在架构方面，原生算力采用数据流架构和脉动阵列等，减少内存搬运，进而提升计算效率；在指令集方面，设计张量指令集以支持矩阵乘加，扩展混合精度与动态量化指令，以适配AI计算特性；在通信协议方面，应用远程直接内存访问（RDMA）等技术实现CPU旁路与直接内存访问等；在精度优化方面，采用FP8低比特训练技术，降低计算与显存开销。

#### 1.4 与配电制冷的协同

AI数据中心（AIDC）相较于传统的互联网数据中心（IDC），除了需要配备完善的IT基础设施外，还需要特别关注电力和制冷系统。这是因为，AI算力需求激增会导致硬件功耗和散热大幅上升。传统IDC主要以通用CPU为主，单机柜功率为4~8 kW，采用风冷方式即可满足散热需求。而AIDC依赖GPU和张量处理器（TPU）集群，单机柜功率通常可达10~100 kW，部分高密度场景甚至更高。高密度算力下硬件易因过热降频或损坏，使得液冷成为了刚需。AI原生的液冷，已经形成了冷板式与浸没式两大技术路径。

### 2 智能原生的业务

各类技术的应用演进一般都会经历原生应用、技术化转型与技术性原生3个阶段。

过去20余年，互联网应用的转型和原生演进过程契合这一规律。在互联网信息展示方面，Yahoo公司的门户网站，本质是传统电话黄页的数字化翻版，用户还需主动逐级检索；Google公司的搜索引擎，则通过算法对网页相关性进行概率排序，以关键词关联链接实现高效触达，原生出了对用户意图的动态理解。在互联网通信方面，IP电话和基于IP的语音传输（VoIP）通过IP协议改造传统语音传输，本质仍是电话服务的数字化延伸；而社交网络基于互联网的即时性与社交属性，原生出了人际交互的新范式，彻底摆脱了传统通信以话音为中心的模式和单纯工具属性。在互联网媒体方面，网络电视（IPTV）以IP网络传输电视信号，但内容生产仍沿袭传统电视的遥控器操作、频道编排和节目播控模式。而如今互联网上的视频服务模式，原生出短剧、竖屏与弹幕互动等的典型特征，这在20多年前讨论三网融合时是难以想象的。

从互联网时代发展到人工智能时代，AI与各行业的融合也将会是一个逐步演进的过程，且大致包括“智能化转型”和“智能原生”两种形态。其中，智能化转型是AI技

术对传统业务的渐进式赋能，其核心目标是通过算法优化并提升现有业务的业务效率、降低成本，相关技术多作为工具嵌入业务流程。智能原生是将AI技术深度融入业务基因后的范式变革，其核心特征是以AI技术为基础（或以AI为核心引擎），实现对现有业务的深度改造，甚至创造出全新业务形态。

当前，已经出现了一些AI原生的新应用。例如，AI编程颠覆了传统手动编码范式，借助自然语言交互实现高效优质的代码生成；硅基人突破了预设脚本的限制，在数字空间中完成高保真孪生建模、动态人格重塑与沉浸式化身创造；AI编剧打破传统内容生产的线性流程，支持主题驱动的自动化剧情生成与实时叙事调整；AI通信并非简单地把AI技术叠加到传统通信之上，而是通过脑机接口等实现思维信号的无损传输与意图直连；AI原生企业则将人工智能深度嵌入企业运营全链路，构建智能决策中枢、持续创新引擎与个性化价值创造网络三大核心能力。

### 3 AI原生代

人是决定技术演进与社会发展的核心因素。计算机和互联网等数字技术把人群分化成数字难民、数字移民和数字原生代等，而AI正在把人群的分化从数字化层面推升级至智能化层面，催生出对应的AI难民、AI移民和AI原生代等。

AI原生代天然会将把互联网和云计算等当作传统基础设施而非新型基础设施。他们将伴随着大模型、智能体（Agent）和具身智能等智能原生新业态共同成长，在社交模式、消费逻辑和生命伦理等方面的理解，即便与传统数字原生代相比，也将呈现出显著差异。当然，AI原生代还必须面对一个全新的挑战，即如何定义碳基生物人与硅基AI人之间的关系。

在社交模式方面，AI原生代将不再局限于线上人与人之间的社交关系，而会进一步扩展到与“AI人”（也称数字人、数智人、数字分身、数字化身、硅基人等）打交道、当朋友、谈生意甚至谈恋爱等。互联网实现了“信息在线”，转移了数字原生代的注意力，并且发展出了一套注意力生态；AI人将实现“情感在线”，会转移AI原生代的情感依托，并且发展出情绪价值的新业态。未来，硅基人很有可能会超越传统社交网络中的碳基人类，为AI原生代和AI移民提供情感陪伴、会话聊天、情感支持甚至参与决策。这将对现代社会的社会伦理秩序和婚姻家庭制度带来全新挑战。

在消费逻辑方面，随着全球物质生产的极大丰富，人类社会正在从以物质消费为核心，过渡到以精神消费为重点的新消费时代。AI原生代尤为关注情绪价值的消费，或者说

属于“情绪价值优先”的一代人。AI原生代将通过算法筛选和推荐，寻找能够满足自身高阶情感需求的AI产品和服务，例如为虚拟偶像消费、购买具有情感共鸣的数字藏品等，推动情绪经济向规模化、专业化方向发展。

从自我认知方面，基于AI的数字分身会使得原生代的身份认同变得更加复杂。传统数字原生代通常随身携带一部智能手机，将其作为连接世界的窗口，以满足信息获取、社交互动和娱乐消费等需求。AI原生代通常会配备一个AI助理（物理形态不限于手机，可以是各类可穿戴设备），为其提供主动服务和智能决策。早期的数字移民生活在一个线上线下相分离的世界，传统数字原生代生活在一个线上线下相结合的世界，而AI原生代将生活在一个线上线下深度融合、真实与虚假共存、碳基智能与硅基智能并列的新世界里。

#### 4 结束语

本文从技术、业务和人3个角度对AI原生展开初步探讨，希望能够对AI的应用落地和模式创新起到一定的参考作用。另外，与互联网应用和云原生概念类似，对智能原生的理解和实践也将是一个不断发展演进和丰富的过程。

#### 参考文献

- [1] IDC. 全球数字化转型支出指南 [EB/OL]. [2025-11-05]. [https://my.idc.com/getdoc.jsp?containerId=IDC\\_P32575](https://my.idc.com/getdoc.jsp?containerId=IDC_P32575)
- [2] 何宝宏. 数字原生 [M]. 北京: 中译出版社, 2023
- [3] 中国政府网. 国务院关于深入实施“人工智能+”行动的意见 [EB/OL]. (2025-08-25) [2025-11-05]. [https://www.gov.cn/zhengce/zhengceku/202508/content\\_7037862.htm](https://www.gov.cn/zhengce/zhengceku/202508/content_7037862.htm)
- [4] Economist. AI firms will soon exhaust most of the internet's data [EB/OL]. (2024-07-27) [2025-11-05]. <https://www.economist.com>
- [5] LI K, CHEN Y, VIEGAS F, et al. When bad data leads to good models [EB/OL]. (2025-05-07) [2025-11-05]. <https://www.arxiv.org/abs/2505.04741>

#### 作者简介



何宝宏，中国信息通信研究院云计算与大数据研究所所长；长期从事互联网、云计算、大数据和人工智能等领域的技术研发、标准制定与产业研究工作。