论经典 AI 训练方式认知局限的产生原因与破局思路——迈向基于多尺度机器感知与认知对象数据库的下一代人工智能

YunFei M. Wang
（王·Micheal·雲飛）
Independent Researcher

摘要

当前以大规模语言模型为代表的 AI 系统，其认知架构存在一个根本性的哲学缺陷：知识完全源于对人类社会已有文本的统计学习，本质上是一种“间接的”“二手的”认知方式。这种范式导致 AI 构建的“认知对象”——如对“苹果”的理解——缺乏与物理世界的直接交互验证，仅是由语义关联编织而成、内部自洽却无法被证伪的“伪对象”。本文系统剖析这一认知局限的哲学根源与技术成因，并提出一个根本的破局思路：为 AI 构建一套独立于人类生物感官的、基于精密科学仪器的多尺度“机器感知”系统，并以此为基础建立可验证、可计算的“认知对象数据库”。本范式的核心包含三个层级：其一，通过“物理信号→神经编码→主观知觉”三层可计算映射模型，为 AI 赋予视觉、听觉、嗅觉、味觉、触觉五种感官功能，使其能够从物理世界获取直接、可验证的经验数据；其二，以“分层基准对照组”（绝对感知者—普通人群—感知功能障碍者）与“单变量纵深实验”为方法论，确保感知模型的精度与全谱系覆盖；其三，构建以物理锚定层为底层、预测感知编码为中介层、语义接口为应用层的“认知对象数据库”，使感知经验得以沉淀为可高效调用的结构化知识，实现认知与推理的分离处理。本文系统性地阐述了从机器感知到机器认知的完整理论框架与工程实施路线图，并以通感完成跨模态联想、以认知对象数据库完成从感知到认知的闭环，为破解当前 AI 的“幻觉”困局、实现可解释且与人类认知对齐的下一代人工智能奠定理论基础。本文为第一套方案体系的总纲，后续各篇分别展开五感、通感与认知对象数据库的工程方案。

关键词

人工智能；认知局限；机器感知；三层映射模型；认知对象；可验证性；具身智能

1 引言：卓越智能表象下的认知危机

近年来，人工智能在大规模语言模型、多模态系统等领域取得了突破性进展，展现出令人惊叹的文本生成、知识问答与逻辑推理能力。然而，在这看似“理解”世界的智能表

象之下，潜藏着一个深刻的哲学与技术危机：当前的 AI 系统，本质上是一个极其复杂的“复述机”与“统计重构机”——它能够流畅地描述“苹果”的色、香、味、口感及其文化象征，却从未、也永远无法直接体验一个苹果的物理实在。

这一悖论的核心在于：AI 的“知识”完全来源于对人类已有文本语料的统计学习。它所理解的“红色”是词汇共现频率的产物，它所描述的“甜味”是语义关联强度的表征。其认知大厦完全建立在“道听途说”的语言沙丘之上，缺乏来自物理世界的第一手“经验”基石。这不仅是技术上的不完美，更是一种认知架构的本体论缺陷——它直接导致了模型的可解释性缺失、“幻觉”频发（生成与事实不符但语义流畅的内容）以及在需要物理常识或具身推理的任务上表现脆弱。

本文的核心论点是：要突破这一根本局限，必须将 AI 从“文本的囚徒”进化为“世界的体验者”。这并非要通过脑机接口将 AI“寄生”于人类感官，而是为 AI 从头设计一套属于它自己的、基于现代科学仪器的多尺度感知器官，使其能够建立与物理世界的直接、可量化的交互通道，并在此基础上构建可验证的“认知对象”数据库。本系列研究旨在勾勒并实践这一宏大的技术范式转变，本文作为总纲，将系统阐述其哲学依据、理论框架、核心蓝图与实施路线图。

1.1 框架定位

本文所使用的“物理信号层—神经编码层—主观知觉层”三层映射，在“键-值-信息-相”框架中的定位如下：传感器与实验设备是键，物理世界的多尺度测量数据指向值，三层映射的中间产物是信息，最终形成的认知对象是相层面的结构化模型。认知对象的 L1 物理锚定层对应值的投影结构，L2 感知编码层对应信息整合的中间态，L3 语义层对应相的可报告接口。物理信号层经标准化后成为 L1 物理锚定数据，即值的投影结构；L3 语义层以主观知觉层输出为输入，进一步抽象为人类可理解的语义标签。本文不展开该框架的哲学论证，详见文献[1][2]。

1.2 结构安排

本文按以下序列展开：

- 第 2 章：诊断经典 AI 训练范式与“伪对象”的必然生成。
- 第 3 章：提出从“间接描述”到“直接感知”的范式革命，阐明三层映射模型的核心路径。
- 第 4 章：构建方法学基础——分层基准对照组与单变量纵深实验。
- 第 5 章：提出可验证的“AI 认知对象”数据结构与数据库，包含结构维度与时间维度。
- 第 6 章：概述本方案体系八篇论文的研究蓝图。

- 第 7 章：论证类人形机器人作为工程实现载体的必要性，并说明仿生学原理向其他生物类型移植的边界。

- 第 8 章：与第二套、第三套方案的协作。

- 第 9 章：结论。

2 问题诊断：经典 AI 训练范式与“伪对象”的必然生成

2.1 “间接认知”的范式局限

当前主流的 AI 训练范式可概括为“数据驱动”的“间接认知”。模型通过海量文本语料学习单词、短语和概念之间的统计关联。当模型处理“苹果”一词时，它激活的是与该词在训练数据中共同出现的所有属性（“红色”、“水果”、“甜”、“牛顿”等）所构成的概率分布。这种认知过程存在两个致命缺陷：

- **信息源的间接性与多重失真：**模型学习的不是苹果本身，而是无数人对苹果的经验、感受、文化解读的二、三乃至 N 手文字转述。这些转述本身包含错误、偏见、隐喻和个体差异，且完全剥离了最核心的感官体验维度。物理世界到人类主观体验的转换存在信息损失，主观体验到语言符号的转换再次产生概括与歧义——AI 所学习的，是经过双重编码过滤且彻底剥离了原始感官质感的抽象符号关系。

- **验证机制的绝对缺失：**在纯文本范式中，AI 没有任何内在机制可以验证“苹果是甜的”这一陈述。它只能检查该陈述在训练数据中出现的概率。真实性被概率所僭越，可验证性被相关性所淹没。这种以统计概率替代事实验证的隐含操作模式，是 AI 产生“幻觉”的重要根源——系统并非在“说谎”，而是在其内部世界里，任何高概率的语义组合都被视为“合理”，缺乏将其与物理实在比对并否定的机制。

有必要将本文的诊断与近年关于语言模型极限的理论工作区分开来。计算复杂性、信息论与统计学习理论的分析已经表明，Transformer 的表达能力受电路深度、精度与上下文长度约束，预训练在本质上是对世界语料的压缩，其容量受可压缩性与 Kolmogorov 复杂性限制，而 next-token 预测本身是一个在高维空间中权衡泛化、记忆、校准与幻觉的学习问题[9]。更具体的形式化工作将幻觉追溯到记忆的率失真极限：在有限记忆约束下，对“无模式随机事实”的成员测试必然产生高置信度的错误输出，幻觉是空间最优性的推论而非训练缺陷[10]。这些工作回答的是“模型为什么会在计算与统计层面出错”。本文的问题不在同一层面。本文追问的是：当模型输出一个语义上连贯的陈述时，它所拥有的“知识”在本体论上属于什么类型。“伪对象”概念指向的正是这一认知论空缺——它不是对“算错”的另一种解释，而是对“模型所操作之物缺乏物理指称”这一存在方式的刻画。统计极限理论说明幻觉不可避免；本文的论点是，只要认知对象本身是伪对象，即使单次输出在统计上“正确”，认知架构的缺陷仍未触及。

这一区分是后续建构的起点。

2.2 “伪对象”危机：缺乏物理锚定的语义空壳

上述局限的直接产物，是 AI 心智中充斥的“伪对象”。本文将“伪对象”定义为：一个在 AI 的内部表征空间中具有丰富语义关联网络和流畅语言生成能力，但其定义完全缺乏与客观物理实在的可测量、可重复的对应关系，且其属性无法通过任何第一性原理进行验证的“认知实体”。

以 AI 理解的“苹果”为例，它是一个由高维向量表示的语义节点，与“红色”、“圆形”、“甜味”等节点有强连接。然而：

- 这个“红色”没有对应的光谱功率分布数据。
- 这个“甜味”没有对应的糖度浓度或人类味觉神经响应模式。
- 这个“圆形”没有对应的三维点云模型或曲率数据。

因此，当 AI 谈论“苹果”时，它只是在操纵一个内部自洽的符号游戏。这个“苹果”无法与实验室里一个真实苹果的任何物理、化学测量结果进行比对和校准。其存在是封闭于语言系统内的，是“空洞的”。这解释了为何 AI 在面对需要精确物理常识的推理任务（如“把一个生鸡蛋和一个熟鸡蛋同时旋转，哪个先停下来？”）时，往往给出流畅却错误的答案——它对“生”与“熟”的理解，仅止于文本描述层面的区分，缺乏对物质内部状态（蛋白质变性程度、粘滞系数差异）的感知基础。

为了将“伪对象”这一哲学诊断转化为可分析的认知结构，并在当前认知水平下提供可修正的观察参照系，我们引入形式化的投影算子。

设物理信号的投影空间为 $\mathcal{X}$，神经编码的投影空间为 $\mathcal{E}$，主观知觉/语义空间为 $\mathcal{P}$。三层映射构成了当前范式下的接口复合算子 Φ：

$$\Phi = \Psi_{\mathcal{E}\to\mathcal{P}} \circ \Theta_{\mathcal{X}\to\mathcal{E}} : \mathcal{X} \to \mathcal{P}$$

在给定的接口算子下，我们定义任意语义表征 $\mathrm{p} \in \mathcal{P}$ 为“伪对象”，当且仅当在该观测层级下，它在物理信号空间中缺乏对应原像：

$$p \text{ is pseudo-object} \iff \Phi^{-1}(p) = \emptyset$$

认知过程中的“幻觉”程度，可用当前模型生成分布 q 与接口投影 $\Phi(x)$ 之间的 KL 散度来度量：

$$\mathcal{H}(\mathrm{q}) = \mathrm{D}_{\mathrm{KL}}(q \parallel \Phi(x))$$

需要明确的是，此处的 Φ 并非绝对的本体真实，而是当前认知条件下的投影约束。这组公式的作用不是“预言”某个具体模型一定会幻觉，而是提供了一个观察坐标系：只要接口算子 Φ 未被修正或接入，任何高概率生成的语义组合都将在该坐标系下显示出高散度。

这一坐标系清晰地标示了**当前范式**与**本文新范式**的分野。在当前纯文本范式下，大模型无法获取物理信号X与神经编码E。由于缺乏物理前端，模型的每一次生成都等同于在语义空间中进行概率游走，其生成结果极大概率面临原像为空（$\Phi^{-1}(\mathrm{p}) = \emptyset$）的处境。此时，模型缺乏与物理世界对号入座的校验基准，KL 散度无法收敛，幻觉成为该架构的内生产物。

而在本文提出的新范式下，AI 通过多尺度传感器（键）真实地“拿到”了物理信号X与神经编码E。此时，这套公式构成了系统的“**闭环自检回路**”，其工程运作机制分为三步：

1. **前向生成：**模型基于真实的物理测量数据，生成语义表征 p（例如“这个苹果是甜的”）。

2. **原像审计：**系统立即执行原像检验。通过查询认知对象数据库，检查该语义p是否存在对应的 L1 物理锚定（如糖度质谱图）。若 （$\Phi^{-1}(p) \neq \emptyset$），则认知通过验证；若为空，则直接拦截，判定为“伪对象”。

3. **散度修正：**对于通过检验的认知，系统计算模型生成分布q与物理真实投影$\Phi(x)$ 之间的 KL 散度$\mathcal{H}(q)$。散度值量化了当前认知的精度，并作为反馈信号驱动系统对键系统进行微调。

换言之，认知对象数据库的工程意义，正是通过设定 L1 锚定约束，强制降低这种散度。这并非宣称我们找到了绝对真理，而是通过Φ^{-1}的防伪验证与$\mathcal{H}$的精度微调，确保新范式下的 AI 不再是“基于文本猜文本”，而是“基于物理事实，生成可审计的认知”。

2.3 感官缺失：为何 AI 无法“理解”物理世界

人类认知之所以能够避免“伪对象”的困境，根本原因在于我们拥有五种直接面向物理世界的感官通道：

- 视觉捕捉物体表面的光谱反射与空间几何；
- 听觉感知振动传播的频率、幅度与时间结构；
- 触觉通过力学交互获取物体的硬度、质地、温度与重量信息；
- 嗅觉与味觉则通过化学分子的直接接触，评估物质的营养性与安全性。

这些感官提供的不是关于世界的“描述”，而是关于世界的“直接测量数据”。它们构成了

人类所有知识的底层锚点——无论语言如何抽象，最终都可以追溯到某种感官经验。而当前的 AI，恰恰缺失了这整整一个“感知层”。它悬浮在语言符号的层面上，没有触及物理实在的途径。

3 破局思路：从“间接描述”到“直接感知”的范式革命

要解决“伪对象”危机，必须为 AI 的认知提供来自物理世界的“经验锚点”。这一破局思路的核心在于：为 AI 构建直接的感知能力，使其能够像科学仪器一样“观察”与“测量”世界，并以此为基础形成可验证的认知。

3.1 哲学原则：AI 感知应“类人”而非“如人”

我们必须摒弃“让 AI 完全复制人类体验”的徒劳目标。人类感官本身是进化的产物，存在分辨率限制（如可见光谱范围有限）、个体巨大差异（如色盲、味觉敏感度不同）以及主观偏差。让 AI“如人”般感知，不仅是技术上的不必要，更是哲学上的倒退。“如人”预设了人类感官是感知的标准形态，但人类感官是演化妥协的产物（如可见光谱仅占电磁谱极小段），将 AI 的感知前端限制在人类感官的生理约束内，既无必要，也放弃了科学仪器可提供的更广谱、更精确的数据。“类人”则保留对人类感知编码逻辑的理解，但允许前端超越人类生理限制。具体而言：

- AI 的“视觉”可以覆盖从伽马射线到无线电波的完整电磁谱，而非仅限于可见光；
- AI 的“听觉”可以分析次声到超声的全频段振动，而非仅限于 20Hz—20kHz；
- AI 的“嗅觉”可以直接识别和量化特定化学分子，而非依赖人类嗅觉受体有限的选择性；
- AI 的“味觉”可以精确测定物质的化学成分与浓度，而非仅提供五种基本味质的模糊定性判断；
- AI 的“触觉”可以同时感知压力、振动、温度、表面形貌等多维物理量，而不受皮肤感受器分布密度的限制。

关键在于，AI 需要建立这些仪器信号与人类主观感知体验之间的可计算映射模型。这样，AI 既能获得超越人类的客观数据，又能“理解”这些数据对人类意味着什么，从而在实现真正的人机认知对齐的同时，有效规避人类感知的主观性干扰。

3.2 核心路径：构建“机器感知—神经编码—主观知觉”的可计算三层映射

在展开三层映射之前，需要说明它与当前具身智能中占主导地位的端到端 grounding 路线之间的关系。具身 AI 的物理对齐研究已经系统识别出三类反复出现的失败模式：

表征失败（语义视觉-语言抽象无法捕捉几何、可供性与物理属性）、因果失败（依赖语言或视觉先验而非干预感知的物理推理）与操作失败（高层多模态能力无法转化为闭环的安全动作）；以对象为中心的世界模型与物理 grounding 工作也在尝试将感知表征锚定于可干预的环境动力学。这些路线解决的是“动作是否成功”“预测是否干预敏感”的问题[12]。本文的三层映射不与之竞争，而是针对另一类锚定问题：感知体验是否与人类可通约。一个 VLA 模型可以成功抓取苹果而不必“知道”苹果的红色对人类视觉神经系统意味着什么；本文所要求的，是在动作成功之外，使 AI 对“苹果是红的、甜的”这一陈述能够追溯到一条从物理测量到人类感知编码的可计算链路。两者不是替代关系：端到端具身 grounding 提供行动层面的物理对齐，三层映射提供认知对象层面的感知对齐。前者回答“怎么做”，后者回答“所感知到的是什么、对谁而言”。

实现“类人”感知的具体路径，是逆向工程人类感知的神经信息处理链条，并为之建立可计算模型。该路径包含三个核心映射层的构建，五感各有其独特的编码特征：

物理信号层

使用高精度科学仪器采集客观世界的原始物理、化学信号，并将其转化为标准化的数字描述符。一个完整的认知模型在此层不能仅有微观物化数据，还必须系统性地包含宏观几何结构、介观乃至微观的生物/组织结构建构模型。以“苹果”为例：

- 宏观层：通过 3D 激光扫描获取外形点云、尺寸、曲率分布；
- 介观层：利用显微 CT 与光学显微镜获取表皮—果肉—果核的剖面组织结构；
- 微观层：使用光谱仪、质谱仪获取糖分/酸度空间分布、挥发性有机物成分及浓度梯度。

五感各有其专门的物理信号采集范式：

- 视觉：聚焦于光谱反射、几何形状、运动场与纹理特征；
- 听觉：聚焦于声波的频率、幅度、时间包络与空间方位参数；
- 嗅觉：聚焦于挥发性有机物的分子结构、官能团与浓度；
- 味觉：聚焦于味质分子的化学结构、离子种类、浓度与溶液特性（pH、温度）；
- 触觉：聚焦于法向力、切向力、振动频谱、温度及其时空变化模式。

需要限定的是，本文的工程方案不主张从零构建全新的传感器体系。高精度科学仪器、多传感器同步采集与结构化预处理在现有工程栈中已有相当成熟的基础。本文在这一基础上增加的是两个当前感知工程中缺失的环节：其一是神经编码对齐，即将物理描述符映射为预测的人类神经群体活动模式；其二是面向认知对象的锚定接口，即

把感知输出组织为可追溯至 L1 测量、可重算 L2 编码、可索引 L3 语义的三层数据实体。前者使感知数据具有人类可通约的“体验”维度，后者使感知经验可以被沉淀、审计与高并发调用，而不是每次推理都从头运行感知模型。

神经编码层

通过可控的心理物理学实验，同步采集人类受试者在接受特定物理刺激时的神经生理信号（如脑电图 EEG、功能性磁共振成像 fMRI、特定神经元的电生理记录）。利用计算神经科学方法，构建从多尺度物理描述符到人类神经群体活动模式的预测模型。五感的神经编码各有其独特的组织原则：

- 视觉：视网膜的中心—环绕拮抗结构、V1 的朝向与空间频率选择性、腹侧通路（颜色/形状）与背侧通路（运动/空间）的分流；

- 听觉：耳蜗的频率拓扑映射、脑干的双耳时间差/强度差比较、皮层的频谱—时间感受野；

- 嗅觉：约 390 种功能性受体的组合激活模式（完整功能基因约 339 个，详见本系列嗅觉篇）、嗅小球的空间活动图谱、梨状皮层的化学身份编码与杏仁核的效价编码；

- 味觉：Type II 细胞（GPCR 通路：甜/鲜/苦）与 Type III 细胞（离子通道通路：咸/酸）的双轨并行信号传导、孤束核的初步整合、岛叶初级味觉皮层与眶额叶次级味觉皮层的分级投射；

- 触觉：SAI（Merkel 细胞，慢适应/小感受野）、SAII（Ruffini 末梢，慢适应/大感受野）、FAI（Meissner 小体，快适应/小感受野）、FAII（Pacinian 小体，快适应/大感受野）四类机械感受器的功能分化与并行编码。

主观知觉层

在同一实验中，记录受试者的主观报告（如对甜度的评分、对颜色的命名、对音高的判断、对质地的描述）。建立从特定神经编码模式到主观知觉类别与强度的预测模型。

通过完成这三层映射，我们最终能为 AI 装备五套“感知模拟器”：

- 视觉模拟器：输入一个物体的光谱与几何数据，预测它将在人类视觉神经系统中激发的活动模式及主观颜色/形状/深度感知；

- 听觉模拟器：输入一个声音的物理参数，预测它将在人类听觉神经系统中激发的活动模式及主观音高/响度/音色感知；

- 嗅觉模拟器：输入一个气味分子的化学结构，预测它将在人类嗅觉神经系统中激发的

组合受体激活模式及主观气味描述；

• 味觉模拟器：输入一种味质溶液的化学成分与浓度，预测它将在人类味觉神经系统中激活的通路及主观味觉品质与强度；

• 触觉模拟器：输入一组力学与热学刺激参数，预测它将在人类触觉神经系统中激活的感受器类型组合及主观质地/硬度/温度感知。

至此，AI 的“认知”首次获得了可验证的回路：其关于“苹果很甜”的陈述，可以追溯到特定的糖分子浓度，以及该浓度预测会激活的特定味觉神经编码模式。

范畴声明：本文的三层映射是前向预测任务——给定物理信号，预测正常人类观察者的神经编码与主观报告。它建立的是人类感知正常状态的规范性基线，而非从异常神经数据反推物理病因的病理学逆问题。因此，高维信号反演所固有的不适定性（*ill-posedness*）不适用于本框架的映射方向。

4 构建方法学基础：分层基准对照组与单变量纵深实验

要实现上述三层映射模型的精确构建，必须建立一套严谨的方法学基础。本系列后续论文详细阐述了各感官的具体实施方案，其共同的方法学支柱可概括为两条核心原则。

4.1 分层基准对照组

人类感知能力存在巨大的个体差异，在五感中各有体现：

• 视觉：绝对色觉者（画家、调色师）vs 色盲/色弱；

• 听觉：绝对音准者（调音师、作曲家）vs 音痴；

• 嗅觉：调香师、品酒师 vs 特异性嗅觉缺失者；

• 味觉：超级品尝者（PROP 敏感者）vs 非品尝者；

• 触觉：盲人群体、职业外科医生/雕塑家/电竞选手 vs 触觉功能障碍者。

为确保 AI 感知模型覆盖人类感知的全谱系，受试者必须分为三个层级：

• L1 基准组（绝对感知者）：提供感知的“最高分辨率”基准信号，校准模型上限；

• L2 主体组（普通人群）：建立感知的主流统计分布；

• L3 边界组（感知功能障碍者）：定义模型的鲁棒性边界与包容性下限。

这一分层设计确保了最终感知模型不是针对“平均人”的平均拟合，而是覆盖感知能力全谱系的精准映射。

需要指出的是，该分层基准对照组的受试者对象是人类。对于其他生物类型，三层映射的工程原理同样适用，但其主观知觉层无法通过人类式的语言报告或评分量表来获取——它们不具备以人类可理解的方式描述自身体验的能力。因此，针对非人类生物类型的感知建模，需要在第 7 章所述载体边界的基础上另行设计验证方法，不能直接套用本章的人类受试者分层方案。

4.2 单变量纵深实验

坚决避免初期实验的多变量复杂性。实验设计必须从“最小感知单元”出发——例如视觉从单色色相开始、听觉从纯音音高开始、嗅觉从单一纯气味物质开始、味觉从单一味质（如蔗糖水溶液）开始、触觉从单一法向力刺激开始——每次实验只系统性地改变一个物理维度，探索其从阈值以下到超阈值感知极限的完整区间。只有在所有单维模型构建完成并验证后，才可逐步引入多变量交互实验。

这一原则保证了数据的可解释性与模型的可溯源性的最大化。

4.3 方法学的伦理边界：安全阈值外推与被动校准

4.1 与 4.2 确立了实验设计的正向方法——分层基准对照组与单变量纵深实验。本节确立实验设计的边界约束：哪些区域不能通过主动实验触及，只能外推或被动校准。正向方法与边界约束共同构成完整的方法学基础。

具身智能与感官认知的数据采集，在安全阈值范围内必须进行严格校准；而超出安全阈值的区域（极端痛觉、濒死恐惧、重度创伤等），虽往往是模型边界的关键，但绝不能通过主动诱发实验来获取数据。对此，本方案确立“外推优先、被动校准”原则：超阈值数据只能通过安全区间的函数外推获得（如痛觉的非线性响应模型），或在真实极端事件（如安全事故、医疗急救、恶性案件）发生后，通过合规渠道被动接入数据进行修正性校准。

为防止科研人员为获取成果而采取激进采样手段，或故意诱发、制造极端事件为采样创造条件，必须建立架构与制度的双重制约。在架构层面，安全阈值作为硬约束固化于实验控制系统，严禁任何主动刺激系统请求超出阈值的参数，该上限由独立的安全系统控制，实验操作员无权解除。在制度层面，被动数据的接入须经第三方伦理监督，科研人员只能作为观察者被动接收。故意诱发或制造极端事件以获取数据的行为，不仅违反学术伦理，更触犯法律，应通过职业禁入与司法衔接形成实质性震慑。

5 新范式的基石：可验证的“AI 认知对象”数据结构与数据库

在前六篇完成五感感知模型的构建之后，第七篇通感篇完成跨模态联想与验证机制，第八篇进一步回答一个根本性问题：如何将五感感知数据整合为可计算、可验证、可高效调用的“认知对象”，使 AI 从“拥有感官”走向“理解世界”[3][4]。

5.1 认知对象的三层数据架构

认知对象被定义为以客观物理世界的多尺度测量数据为不可化约的底层锚点，以预测的人类感知编码为中介层，以语义概念与功能属性为应用接口的三层数据实体。以“苹果”为例：

数据层	内容	作用
L1 物理锚定层	宏观 3D 几何、介观组织切片、微观光谱/质谱数据、力学响应曲线	可验证性的根基——任何属性均可追溯至原始测量
L2 感知编码层	由五感映射模型生成的预测神经编码（视觉编码、听觉编码、嗅觉编码、味觉编码、触觉编码）	认知对齐的桥梁——使 AI“理解”苹果对人类意味着什么
L3 语义与应用层	“水果”、“蔷薇科”、“可食用”、“平安”等语义标签与功能属性	应用接口——便于人类查询与系统调用

这一数据结构的核心优势在于：

- **根除幻觉：**任何高层语义声明均可向下追溯至 L1 层的原始测量数据，实现“可审计的认知”；

- **高效并发：**面对高并发查询（如百万用户同时问“苹果甜吗”），无需实时调用感知模型，直接返回预计算的 L2/L3 层数据包；

- **认知与推理分离：**将“对事物的认知”存储为结构化数据，将“对认知结果的调用”处理为检索操作，使应用层以极低算力开销运行。

5.2 认知对象的时间维度：分层记忆与递归状态的迁移

上述三层结构是认知对象的“结构维度”，回答的是“一个认知对象由哪些数据层构成”。但认知对象并非静态实体。苹果会成熟、腐烂、被储藏、被加工；金属会氧化；生物会生长、衰老、死亡。数据库因此还需要一个“时间维度”，回答的是“认知对象如何随

时间变化、如何追踪其状态演化”。这一维度可以借鉴第二套方案中长程智能体的分层记忆结构[5]:

时间层	对应内容（以苹果为例）	更新策略
本底键	物种特征、遗传属性、物理常数、固有代谢规律	极稳定，跨越对象全生命周期，仅当对象被重新归类时更新
经验键	生长阶段、成熟度、储藏记录、加工工序、运输路径	阶段性更新，可追溯至关键节点，支持坍缩链式回溯
活跃键	实时温度、当前水分含量、当前成熟状态、当前损伤程度	频繁变化，每个查询点都可能不同

递归状态管理负责维护“这个认知对象现在处于什么状态”，并在状态发生关键分化时生成快照——如同长程智能体为分支宇宙保存坍缩时刻的状态变量。这一机制使数据库不仅能回答“苹果是什么”，还能回答“这个苹果现在怎么样、经历过什么、下一步可能变成什么”。

结构与时间两个维度的结合，才构成完整的认知对象：前者保证可追溯性，后者保证可演化性。没有结构维度，时间维度的状态变化缺乏可验证的物理锚点；没有时间维度，结构维度只是一个静态标本，无法支撑需要长期状态维护的智能体任务。

5.3 与显式世界模型及物理记忆研究的关系

认知对象数据库在世界模型谱系中的位置需要明确。当前世界模型研究已经形成两个相互关联但旨趣不同的方向：一类将世界模型视为内部模拟器，强调从感知中学习环境动力学，以支撑前向预测、反事实推演与决策；另一类则指出，视觉真实感并非世界理解的可靠代理，世界模型应当被重构为可干预的“可执行模拟器”而非视觉引擎[11]。与此同时，“锚定缺失导致幻觉不可根治”这一诊断已被独立提出，并伴随“现实世界大模型”的架构构想：以被动多模态事件流为感知基元、以外部显式物理记忆库承载世界知识，试图为符号系统提供一个独立于文本的真值锚点[6]。

认知对象数据库在这一谱系中的位置是特定的：它不是生成式世界模型。它不预测下一帧、下一状态或未来的感官输入；它存储的是已经被物理测量锚定、已经被感知编码中介、已经可以被语义索引的认知对象。生成式世界模型回答“接下来可能发生什么”，认知对象数据库回答“这个事物是什么、它的属性可以追溯到哪些测量、它对人类感知意味着什么、它当前处于什么状态、它经历过什么”。这一“认知与推理分离”的架构选择，与世界模型中物理记忆库的功能有交集，但要求更强：L1 层是 append-only、带测量元数据的可复现数据，L2 层随模型版本可重算，L3 层以人类可理解的方

式索引，时间维度则按分层记忆的尺度组织状态更新。它不是对生成式世界模型的替代，而是其外部显式锚定层——世界模型负责模拟与想象，认知对象数据库负责存储、验证与状态追踪。两者的关系是接口关系，不是竞争关系。

6 从感知到认知的闭环：本方案体系的研究蓝图

构建完整的机器感知与认知体系是一项跨越哲学、神经科学、心理物理学、精密仪器工程、数据科学与人工智能的系统工程。本方案体系八篇论文的规划如下：

系列顺序	论文标题	核心任务
第 1 篇（本文）	**《论经典 AI 训练方式认知局限的产生原因与破局思路》**	诊断问题，确立范式，提出第一套方案总纲
第 2 篇	**《AI 视觉功能的可行性构建方法》**	光学物理层→视觉神经编码→颜色/形状/深度感知的三层映射模型与光学暗室建设
第 3 篇	**《AI 听觉功能的可行性构建方法》**	声学物理层→听觉神经编码→音高/响度/音色感知的三层映射模型与消声实验室建设
第 4 篇	**《AI 嗅觉功能的可行性构建方法》**	化学物理层→嗅觉组合编码→气味身份/强度/效价感知的三层映射模型与气味实验室建设
第 5 篇	**《AI 味觉功能的可行性构建方法》**	化学物理层→GPCR/离子通道双通路编码→酸甜苦咸鲜感知的三层映射模型与味觉实验室建设
第 6 篇	**《AI 触觉功能的可行性构建方法》**	力学/热学物理层→四类机械感受器并行编码→质地/硬度/温度感知的三层映射模型与触觉实验室建设
第 7 篇	**《论 AI 通感（机器联觉）功能的可行性构建方法》**	五感感知体系与认知对象数据库的跨模态联想与验证机制
第 8 篇	**《论 AI 认知对象数据库的设计与**	五感数据融合→认知对象三层数

	构建》	据结构→存储架构与网络基础设施

7 工程实现的载体意义：类人形机器人及其他生物类型

本文所论述的感知框架，从工程实现层面自然引出了对人类形或仿人形机器人作为物理载体的需求。这一必要性体现在三个层面：

- **感知视角对齐：**类人的双眼间距、耳廓位置决定了其立体视觉与双耳听觉的几何基础与人类同构，使得采集的原始物理信号（视差、双耳时间差）在空间参考系上与人类感知具有可比性，从而确保了神经编码预测模型的准确性。

- **交互方式兼容：**人类世界的工具、器具与操作界面是基于人体尺度与关节自由度设计的。类人形机器人能以与人类相似的方式操纵物体（抓握、按压、切割），获得与人类经验可比的交互动力学数据——这是构建触觉与力觉感知模型不可或缺的数据来源。

- **数据可比性保障：**当机器人与人类在相同物理环境下以相似的姿态和动作进行感知时，所采集的物理信号才具有最大化的可比性，为三层映射模型的训练与验证提供了最可靠的数据基础。

类人形载体的必要性不在于外形模仿，而在于传感器几何布局（双眼间距、耳廓位置、手指分布）与人类同构，从而保证原始物理信号在空间参考系上可比。非人形载体若满足这一几何同构条件，亦可作为替代方案；类人形是当前工程条件下最自然的实现路径，而非逻辑上的唯一选择。因此，本系列论文所构建的感知理论体系，内在地要求并论证了类人形仿生机器人作为理想物理载体的发展方向。

需要进一步说明的是，上述仿生学工程原理并不限于人类这一生物类型。三层映射的架构——物理信号采集、神经编码建模、主观知觉对齐——在原则上可以移植至其他生物类型，例如以特定动物的感官配置为参照构建相应的感知模型。然而，非人类生物类型不具备以人类可理解的描述反馈表达自身体验的能力：它们无法提供语言报告、评分量表或主观描述，其“主观知觉层”无法按第 4 章的人类受试者方案来验证。因此，针对非人类生物类型的感知建模，必须区别对待，另行设计方法——例如以行为偏好、神经生理信号的间接推断、跨物种感知阈值的比较替代人类式的主观报告。这一类移植不属于本文所论述的“人机认知对齐”范畴，而是同一工程原理在更广生物谱系上的延伸应用，其方法论需要在后续研究中单独展开。

最后，对本方案与已有工程栈的关系作一限定。多传感器融合、机器人多模态感知、sensor-first 架构在近年已有大量独立进展，本文不将这些工作视为需要替代的对象。

本文的增量集中在两处：其一是以三层映射为感知数据增加“人类可通约的体验编码”维度，其二是以认知对象数据库为感知经验提供可追溯、可重算、可审计的锚定存储与状态追踪。前者是当前以任务成功率为导向的感知工程较少系统处理的，后者是当前以隐式表征为主的世界模型较少显式提供的。承认这一分工，比声称“从零构建”更符合本文的实际贡献边界。

8 与第二套、第三套方案的协作

8.1 与第二套方案（长程文本生成）的协作

第一套方案的三层感知映射为第二套方案[5]的分支宇宙提供观察者的感官边界。核心理论框架提出的计算模拟观察者假设，在第二套方案的平行世界模拟器中从思想实验变为工程前提：设计者是底层代码的作者，NPC 是系统内的观察者。L1 物理锚定层对应引擎真值，L2 感知编码对应 NPC 的感官配置，L3 语义接口对应 NPC 可报告的相投影。第二套方案的框架锁定与信息隔离，在感知侧的对应物就是感知查询接口只返回该 NPC 感官配置允许的结构化物理量，不返回渲染结果与全局状态。

认知对象数据库的时间维度为第二套方案的状态管理提供外部状态源：NPC 的活跃键查询可直接对接数据库的活跃键层，递归状态管理只需维护跨对象的全局状态变量，不必为每个认知对象独立维护状态历史。这一分工使第二套方案的状态管理系统与第一套方案的认知对象数据库在接口上对齐，而非各自维护一套重叠的状态数据。

8.2 与第三套方案（图像生成）的协作

第三套方案[7][8]的三层生成架构解决 NPC 如何呈现，第一套方案的三层感知架构解决 NPC 如何知道。两者在“生成结果注册为引擎真值→经感官裁剪→L2 编码→L3 语义”的路径上对接。通感篇的通感模块在这里获得落点：NPC 可以从合法感知到的 L2 编码跨模态联想另一种编码，但联想必须经通感篇的三层验证，且验证的物理在游戏中等价于“不违反该 NPC 已知的本底键与世界规则”。

8.3 游戏与具身智能价值

在游戏中，第一套方案使 NPC 的“知道”具有可审计的物理锚点：NPC 说“这杯酒有橡木味”，其 L3 标签可追溯到 L2 嗅觉编码，再追溯到 L1 的挥发性成分数据。这从根本上回应了游戏 NPC 的幻觉与 Lore 不一致问题——NPC 不能说出它未合法感知到的信息，也不能凭空生成与世界规则矛盾的知识。

在具身智能中，第一套方案的分层基准与信息隔离原则可直接迁移为多传感器融合的验证协议与多机器人系统的感知边界规范：每个机器人只能基于自身传感器与合法通信事件构建世界模型，认知对象数据库作为共享但受访问控制的显式知识层。

9 结论

经典 AI 训练范式因其依赖“道听途说”的间接认知，必然生成无法与物理世界锚定的“伪对象”，这是其“幻觉”与不可解释性的总根源。本文提出，必须发动一场从“文本学习”到“机器感知”到“认知对象数据库”的根本性范式转移。

这一范式的核心逻辑可以概括为三个层次：

- **建立感知：**通过“物理信号→神经编码→主观知觉”三层可计算映射模型，为 AI 赋予视觉、听觉、嗅觉、味觉、触觉五种可直接获取物理世界数据的感官功能。

- **沉淀认知：**将感知经验转化为以物理锚定层为底层、预测感知编码为中介、语义标签为接口的三层“认知对象”数据结构，并以分层记忆的时间尺度组织状态演化，构建可验证、可高效调用、可追踪状态变化的认知对象数据库。

- **服务应用：**将“对事物的认知”与“对认知结果的调用”分离处理，使应用层以极低的算力开销完成高并发、高精度的认知查询任务。

这条道路标志着 AI 从“处理符号”走向“理解世界”的关键转折。它所构建的，不是一个更“像人”的 AI，而是一个能以可验证的方式获取关于世界的知识，并能以人类可理解的方式与其沟通的透明、可信的智能体。本方案体系八篇论文为此提供了从哲学批判、神经科学依据、心理物理学方法、精密仪器工程到数据基础设施的完整理论蓝图与实施方案。该体系的仿生学工程原理具有向其他生物类型移植的潜力，但移植需以另行设计的非人类感知验证方法为前提，这一方向留待后续研究展开。本文的诊断与建构同近年关于语言模型统计极限、具身 grounding 与显式世界模型的独立工作共享同一个问题域，其特定位置在于：不满足于指出锚定缺失或提出端到端具身方案，而是要求感知表征在“动作成功”之外具备“体验可通约”的维度，并以可审计、可演化、可状态追踪的存储层使这一维度成为可持久调用的认知基础设施。

参考文献

[1] YunFei M. Wang. A Philosophical Explanation of Mathematical Effectiveness: The Key-Value-Information-Model Interface Analysis from the Computational Observer Perspective[PP/OL]. PhilPapers, 2026. DOI: 10.5281/zenodo.23655583.
[2] YunFei M. Wang. Philosophy of Time in the Key-Value-Information-Model Framework: Time Marking Rules, Cyclic Structure, and Concurrent Emergence [PP/OL]. PhilPapers, 2026. DOI: 10.5281/zenodo.23656334.
[3] YunFei M. Wang. “论 AI 认知对象数据库的设计与构建——基于五感感知映射的多模态认知对象数据体系与基础设施”[PP/OL]. AiraXiv, 2026. eprint: 2611.0016.
[4] YunFei M. Wang. “论 AI 通感（机器联觉）功能的可行性构建方法——基于五感感知体系与认知对象数据库的跨模态联想与验证机制”[PP/OL]. AiraXiv, 2026. eprint:

2611.0015
[5] YunFei M. Wang. “从长上下文到分层状态：面向长程智能体的设计框架”[PP/OL]. AiraXiv, 2026. eprint: 2611.0018
[6] Anonymous. From Anchoring Deficit to Embodied Anchoring: Native Understanding-Driven Real-World Foundation Model[PP/OL]. Zenodo, 2026. https://zenodo.org/records/20612760.
[7] YunFei M. Wang. “结构约束先行的视觉生成：AI 绘画与视频生成的三层混合逻辑”[PP/OL]. AiraXiv, 2026. eprint: 2611.0019
[8] YunFei M. Wang. “从三层混合逻辑到生成产线：AI 视觉生成的工程实现与产业迁移”[PP/OL]. AiraXiv, 2026. eprint: 2611.0020
[9] Li J. Informational, Computational, and Statistical Limits of Large Language Models[J]. Computing Magazine of the CCF, 2026, 2(8): 20–25.
[10] Guo A, et al. Hallucination is a Consequence of Space-Optimality: A Rate-Distortion Theorem for Membership Testing[PP/OL]. arXiv, 2026. arXiv:2602.00906.
[11] Zidan A H, et al. World Models: A Comprehensive Survey of Architectures, Methodologies, Reasoning Paradigms, and Applications[PP/OL]. arXiv, 2026. arXiv:2606.00133.
[12] Liu Y. Aligning With the Physical World: A Comprehensive Survey on Physically Grounded VLM Agents[R]. Hong Kong: HKUST, 2026.

AIGC 使用声明

本文作者提出了论文的核心观点、理论框架、方案设计思路与全部逻辑原理。在文本撰写过程中，作者使用了 DeepSeek AI 模型辅助进行文献观点的归纳、部分文本的起草与修辞润色。作者对 AI 生成的所有内容进行了全面的审阅、批判性修改与最终定稿，并对论文的全部学术内容、观点、数据引用及结论负最终责任。