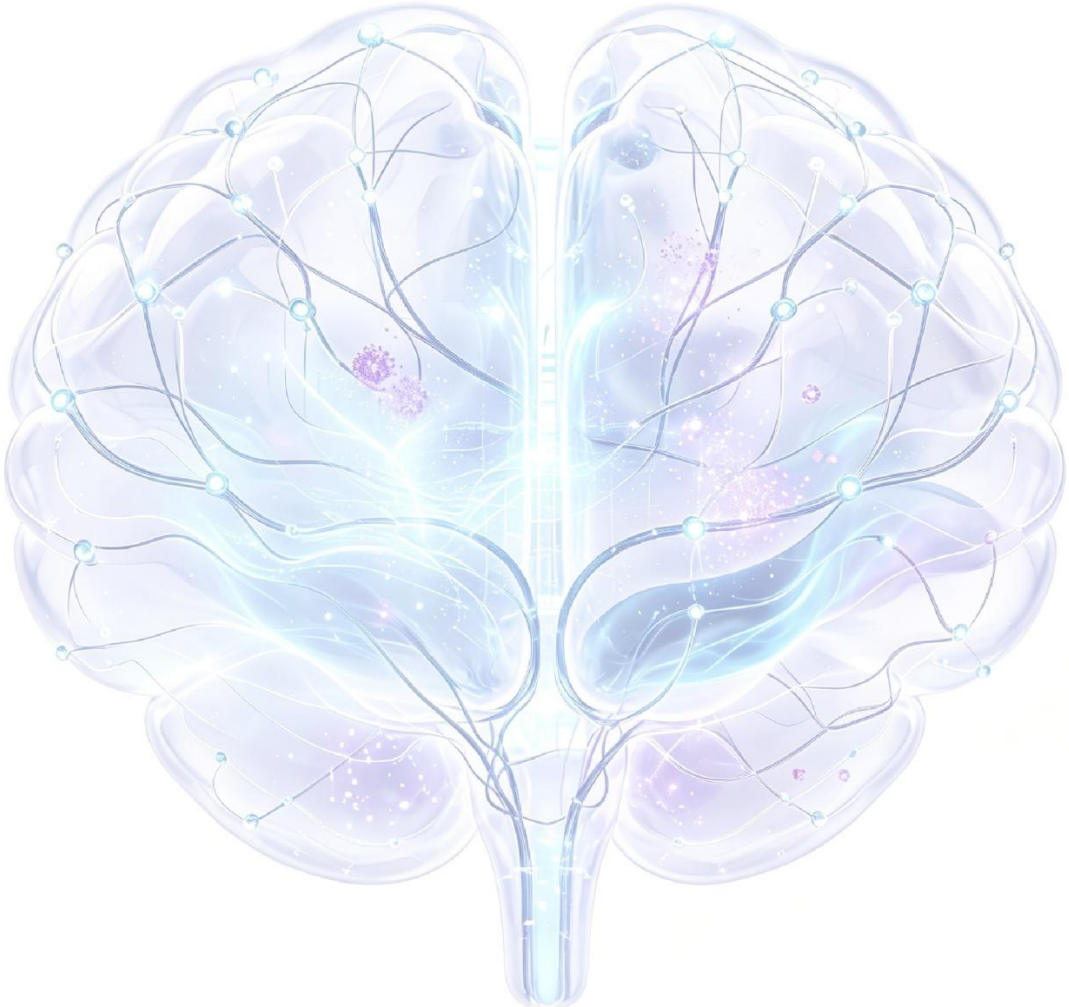
类脑认知世界模型技术白皮书

White Paper on Cognitive World Model Technology

—— 从人脑智能本源到类脑认知世界模型的技术发展探索

— Exploration of Technological Development from the Origin of Human Brain Intelligence to Cognitive World Model



编委会

组织编写

类脑智能产业创新发展联盟

参编单位

复旦大学

中国科学院生物物理研究所

中国科学技术大学

合肥综合性国家科学中心人工智能研究院

类脑智能产业创新发展联盟

上海具脑磐石科技有限公司

智脑工场（上海）智能科技有限公司

上海杨浦科技创新（集团）有限公司

上海长阳创谷企业发展有限公司

版权声明

本白皮书版权属于编委会，并受法律保护。转载、摘编或利用其他方式使用本白皮书文字或者观点的，应注明“来源:《类脑认知世界模型技术白皮书》”。白皮书内容不构成任何投资建议、产品承诺或专利授权。

前言

当前，新一轮科技革命和产业变革加速演进，人工智能（AI）作为引领未来的战略性技术，正深度重构全球产业格局与国家竞争范式。随着 AI 技术从感知智能、认知智能向具身智能延伸拓展，世界模型（World Model）凭借对物理世界的表征、预测与推演能力，被公认为突破大模型范式瓶颈、迈向通用人工智能（AGI）的核心技术底座，成为全球主要经济体科技战略布局的制高点与产业竞争的新赛道。

放眼全球，世界范围内的科技竞争已向底层基础理论与核心技术架构纵深延伸。美国、欧盟等主要经济体纷纷将类脑智能、世界模型纳入国家前沿科技战略与长期研发计划，通过国家级项目持续投入，集聚顶尖科研力量与产业资源加速攻关；全球科技巨头与顶尖学术领袖加速技术布局，从 3D 空间建模、视频生成仿真到抽象表征学习、主动推理框架，多路线并行探索，技术迭代持续提速。一场围绕下一代 AI 底层架构与技术范式的全球竞争已经全面展开，能否抢占世界模型技术先机、构建自主可控的技术体系，直接关系到各国在未来 AI 产业生态中的话语权、规则制定权与核心竞争力。

与此同时，人工智能正从数字空间任务向开放物理环境中的具身交互延伸。现有基础模型与 VLA 在数据效率、长时程闭环控制、开放环境泛化和端侧能效等方面仍面临挑战。莫拉维克悖论描述的是“人类轻易完成的感知运动任务对机器反而困难”等历史观察，并非一个可被单一路线直接“解决”的技术命题。脑科学、机器人学、控制理论、强化学习与因果学习均可为下一代具身智能提供重要启发。

从国内发展来看，我国人工智能产业规模持续壮大，应用场景全球领先，产业体系日趋完善，在世界模型相关的视觉生成、机器人控制、仿真训练等领域已形成一定的技术积累与产业布局。但总体而言，我国仍面临底层原创理论供给不足、高端交叉学科人才匮乏、软硬件协同生态不完善、行业标准与评测体系缺失等挑战。与此同时，我国也具备独特的发展优势与战略机遇：超大规模的应用市场、完备的制造业体系为技术落地提供了广阔场景；科技创新 2030：脑与类脑研究的深入推进，“十五五”规划明确将类脑智能、世界模型作为人工智能领域重点发展方向，政策支持力度持续加大。

为深入贯彻落实国家创新驱动发展战略，助力人工智能高水平科技自立自强，推动类脑认知世界模型产业高质量发展，类脑智能产业创新发展联盟组织具脑磐石、上海类脑智能未来产业集聚区、智脑工场、复旦大学、中科院生物物理所、中国科学技术大学、合肥综合性国家科学中心人工智能研究院等产学研多方专家力量，历时数月深入调研与系统论

证，编制形成本《类脑认知世界模型技术白皮书》。

本白皮书旨在凝聚产业发展共识、明晰技术演进路径、指引企业战略方向，为政府部门制定产业政策、科研机构开展技术攻关、企业布局发展赛道、投资机构判断产业方向提供权威参考与决策支撑。期待与产业各界一道，坚持理论驱动与工程落地并重，坚持自主创新与开放协同并行，走难而正确的技术创新道路，共同推动我国类脑智能相关技术突破与产业繁荣，为培育新质生产力、建设科技强国贡献力量。

目录

前言.....4

第一章 世界模型热潮之下的概念混沌与底层危机..... 9

1.1 当前具身智能范式瓶颈：视觉-语言-动作模型（VLA）的能力边界..... 9

1.2 世界模型（World Model）成为下一代 AI 战略制高点.....10

1.3 行业共性问题：概念模糊与底层机理缺失.....13

1.4 世界模型技术路线全景对比分析.....14

第二章 理论溯源：人脑如何构建自身的“认知世界模型”..... 21

2.1 演进脉络：从哲学猜想到计算科学体系.....21

2.2 神经机制基础：多脑区协同的认知系统.....22

2.3 核心能力维度：世界模型的五大认知支柱.....24

2.4 统一理论框架：主动推理与自由能原理.....25

2.5 脑科学对世界模型的启示.....30

第三章 类脑视角下世界模型的定义、架构与能力边界..... 31

3.1 四类易混淆的“模型”概念.....31

3.2 类脑认知世界模型的核心定义.....32

3.3 类脑认知世界模型五层技术架构.....34

第四章 类脑智能全产业链协同创新发展建议与行动倡议..... 36

4.1 类脑认知世界模型从算法到芯片的产业链路.....36

4.2 专项布局：推动类脑智能重大科技专项落地.....37

4.3 数据基建：建设国家级物理智能开放数据体系.....37

4.4 标准先行：加快行业评测标准与规范体系建设.....37

4.5 开放合作：构建产学研用一体化产业共同体.....38

第五章 总结与展望 39

5.1 核心总结 39

5.2 发展展望 40

5.3 类脑智能根技术创新的战略意义 40

参考文献： 42

图表目录

图 1 世界模型技术流派解析.....14

图 2 自由能原理（Free Energy Principle, FEP）图示.....26

图 3 主动推理（Active Inference）自上而下和自下而上的两套逻辑视角.....27

图 4 智能体基于自由能原则（FEP）的主动推理.....28

图 5 基于变分自由能与期望自由能的主动推理计算体系.....29

图表 1 全球五大世界模型技术流派多维度对比.....18

图表 2 市场上主要的几类“模型”概念对比分析.....31

第一章 世界模型热潮之下的概念混沌与底层危机

本章对世界模型技术演进的分析，从当前具身智能领域的主流实践形态 —— 视觉 - 语言 - 动作模型（VLA）切入。VLA 是大语言模型向物理世界延伸的核心产物，也是现阶段产业界落地“世界模型”概念最普遍的工程载体：当前多数面向具身场景的世界模型方案，本质上都以 VLA 为核心架构，通过大模型串联感知、理解与动作链路，实现对环境的交互与适配。因此，剖析 VLA 的能力边界与底层困境，本质上就是在拆解当前大模型驱动的世界模型的范式天花板；只有先厘清现有技术路径的瓶颈根源，才能准确把握世界模型技术演进的核心动因与突破方向。

1.1 当前具身智能范式瓶颈：视觉-语言-动作模型（VLA）的能力边界

过去十年，以深度学习为核心的第三代人工智能浪潮取得突破性进展，大语言模型（LLM）的爆发推动 AI 从感知智能迈入认知智能阶段。但随着 AI 技术向物理 AI、具身智能、通用智能延伸，以 LLM 为基础的视觉-语言-动作模型（VLA）逐渐显现难以突破的能力边界，集中体现为四大痛点：

- ✧ **小样本学习效率低下**：VLA 依赖海量标注数据的预训练范式，数据成本高、高质量数据稀缺，无法像人类一样通过少量样本快速学习新技能，难以满足真实场景中持续学习、快速适配的需求。
- ✧ **开放环境泛化能力不足**：VLA 模型基于大规模预训练数据的统计拟合，分布内场景表现优异，但在开放、动态、非结构化的真实物理环境中，面对未见过的物体、场景和任务时性能断崖式下降，难以适配具身智能体复杂多变的交互需求。
- ✧ **长时序因果推理薄弱**：VLA 擅长基于统计关联的模式匹配，但缺乏对物理规律、因果逻辑的深度理解，难以进行长时序状态推演与后果预判，无法支撑机器人自主决策、复杂任务规划等具身智能核心需求。
- ✧ **端侧部署算力功耗矛盾突出**：VLA 参数量与计算量呈指数级增长，依赖云端高密度算力集群支撑，而具身智能体通常需要端侧实时响应，高功耗、高时延的云端推理模式无法满足移动机器人、边缘设备的部署要求。**端侧部署需要模型压缩、稀疏化、专用加速、端云协同与实时控制共同优化。**

当前，VLA 是具身智能 1.0 阶段的主流技术方案，其核心逻辑是以大语言模型为中枢，

串联视觉感知、语言理解与动作控制，实现端到端的机器人控制。但 VLA 本质上仍是大语言模型能力向具身场景的延伸，并未从底层突破深度学习的范式局限，整体陷入“高数据依赖、低泛化能力、灾难性遗忘、高运行功耗”的四大困境。

近几年最前沿的技术研发和行业实践已经证明，单纯依靠数据和参数的规模扩张与软件工程优化，无法从根本上解决开放环境具身智能的核心问题。下一代具身智能的技术突破，必须跳出纯计算机科学的工程试错路径，向更底层的生物智能与脑科学寻求理论指导与范式创新。

1.2 世界模型（World Model）成为下一代 AI 战略制高点

面对大模型范式的结构性瓶颈，世界模型（World Model）凭借对物理因果关系的理解、环境状态的动态预测与认知推演能力，被普遍视为突破当前技术瓶颈、迈向通用人工智能与具身智能的核心底座，迅速从学术研究走向产业核心，成为全球科技竞争的新战略高地。

1.2.1 全球顶尖力量集体布局

全球科技巨头与顶尖学术领袖纷纷将世界模型作为下一代 AI 的核心战略方向，多路径并行探索，加速技术迭代：

- ✧ **基础理论派：**世界模型并非人工智能领域凭空诞生的工程概念，其核心思想根植于生物智能的运行机理 —— 生物大脑本质上就是一套高效的内部世界模型系统，通过构建环境的抽象认知表征、预测外部状态变化来指导行为，这是智能体适应开放环境的核心底层逻辑。脑科学与世界模型并非跨领域的简单结合，而是“机理 - 实现”的本源对应：大脑的预测编码机制对应世界模型的表征更新逻辑，海马体认知地图对应世界模型的空间表征体系，主动推理框架对应世界模型的运行闭环，为类脑认知世界模型路线提供了完整的第一性原理支撑。英国理论神经科学家、皇家学会院士 、认知神经科学泰斗 Karl Friston，其提出的自由能原理与主动推理理论，是首个能够完整解释世界模型存在意义与运行逻辑的数学化理论体系，直接催生了主动推理派世界模型技术路线。目前 Friston 的理论已成为类脑认知世界模型的核心理论纲领，其本人也作为全球多家主动推理技术企业的首席科学顾问，推动基础理论向工程化世界模型落地转化，是连接脑科学机理与世界模型技术的核心纽带。此外，图灵奖得主 Yann LeCun 离开 Meta 后创立 AMI Labs，全力推进 JEPA（联

合嵌入预测架构）技术路线，明确提出“世界模型是通往通用人工智能的必经之路”，从抽象表征层面重构 AI 学习范式。

- ✧ **技术延伸派：**OpenAI 推出 Sora 等视频生成模型，向“世界模拟器”方向演进，试图通过生成式 AI 掌握物理世界规律；Google DeepMind 基于仿真环境探索智能体交互学习，通过 Dreamer 系列模型推进基于模型的强化学习；李飞飞创立 World Labs，聚焦 3D 空间智能建模，从空间维度构建世界表征基础。2026 年 9 月，World Labs 发布 Atlas 多模态世界模型，系全球首个从零预训练的全模态空间世界模型，原生融合文本、图像、视频、相机位姿与 3D 深度信息，可在同一套架构内完成像素级相机控制的视效生成、真实场景三维重建与时空连续模拟。
- ✧ **产业落地派：**英伟达、微软、谷歌等科技巨头企业依托算力、平台与生态优势，布局世界模型基础设施与开发工具，推动技术产业化落地。波士顿动力（Boston Dynamics）2026 年推出全电驱量产版 Atlas 人形机器人，搭配谷歌 Gemini 机器人世界模型，可在工厂环境中自主识别环境、规划作业路径、完成负重搬运与工位装配，已在现代汽车蔚山工厂完成多场景商业验证，全年产能订单全部锁定，聚焦汽车制造、重工业等重载场景落地。Figure AI 核心产品 Figure 02 人形机器人搭载自研 Helix 三层架构世界模型系统，联合 OpenAI 强化语义与推理能力，2026 年已进入宝马工厂开展物料搬运小批量试点。全球顶尖力量的集体押注，标志着世界模型已从前沿概念升级为下一代 AI 技术的核心赛道。

1.2.2 国家层面战略持续加码

世界模型是脑科学与人工智能交叉融合的核心技术载体与典型应用场景：脑科学为世界模型提供底层认知机理（包括预测编码、主动推理、海马认知地图、分层记忆结构等），是突破当前大模型范式瓶颈、走向类脑认知世界模型的理论源头；人工智能则为脑科学的理论落地提供算法框架、算力支撑与工程实现路径，二者深度融合推动世界模型从“数据拟合的感知模拟”向“机理驱动的认知推演”升级。

全球主要国家均将脑科学与人工智能融合上升为国家战略，本质上都是通过交叉领域布局抢占下一代世界模型的技术制高点：

- ✧ 美国通过“脑计划（BRAIN Initiative）”持续加码神经科学基础研究，同步由 DARPA、NSF 等机构设立专项，推动脑科学机理向人工智能系统的迁移应用，重点支持基于主动推理、预测编码的世界模型架构，以及面向具身机器人的类脑认知

系统研发。2025 年以来进一步加大对神经形态计算、类脑芯片的资助力度，从硬件层面支撑低功耗世界模型的端侧部署，形成“基础脑科学 - 类脑算法 - 神经硬件 - 场景落地”的全链条国家战略布局；

- ✧ 欧盟“人类脑计划（Human Brain Project, HBP）”于 2023 年 9 月正式结束，项目为期 10 年，汇聚 19 个国家共 155 个研究机构参与。项目周期内产出了高精度数字脑图谱、多尺度脑模拟平台、神经形态计算原型与类脑机器人技术等核心成果，验证了脑科学机理向人工智能系统迁移的可行性，为世界模型的类脑化路线提供了早期理论与技术支撑。项目收官后，核心研究基础设施通过 EBRAINS 2.0 项目（2024-2026 年）持续运营，保留脑建模、认知仿真等公共能力；同时欧盟在“地平线欧洲”框架下设立 2025-2027 年神经技术专项计划，划拨 1.2-1.5 亿欧元资金，重点支持脑科学与类脑智能、具身世界模型的转化研究与标准制定，推动基础研究成果向产业落地延伸。欧盟依托“人类脑计划（Human Brain Project, HBP）”的深厚积累，推动脑科学与人工智能交叉融合，在主动推理、神经形态计算等基础理论领域保持全球领先；
- ✧ 我国高度重视类脑智能与世界模型技术发展，“中国脑计划”以“一体两翼”布局推动脑科学研究与类脑智能转化；国家自然科学基金委 2025 年发布的《关于可解释、可通用的下一代人工智能方法重大研究计划》，明确将“新一代脑启发的人工智能”“类人认知学习框架”列为重点支持方向；“十五五”规划进一步将类脑智能、世界模型作为人工智能领域重点发展方向，从国家层面推动技术攻关与产业培育。上海、北京等地依托区域产业优势，出台专项政策支持类脑智能世界模型的技术研发与产业落地，打造国家级创新集聚区。

1.2.3 产业资本加速涌入赛道

世界模型被视为继大语言模型之后人工智能领域最大的技术变革与产业机遇，全球产业与资本持续加大布局力度。技术突破将带动机器人、智能制造、自动驾驶、数字孪生等多个万亿级产业的智能化升级，催生全新的技术生态与经济增长点。能否抢占世界模型技术先机，直接决定了各国在未来全球 AI 产业格局中的地位与话语权。

1.3 行业共性问题：概念模糊与底层机理缺失

在世界模型快速升温的同时，产业发展也呈现出明显的早期阶段特征：认知偏差、路径依赖与生态割裂等问题交织，制约了技术的系统性突破与产业健康发展。

1.3.1 概念定义模糊泛化，认知边界模糊

当前行业对“世界模型”缺乏统一的定义与标准，概念模糊泛化。视频生成、3D 渲染、物理仿真、游戏引擎等技术形态，都被冠以“世界模型”的名义进行传播与推广，甚至将视觉效果等同于世界模型能力。这种认知混淆的背后，既有技术发展早期定义逐步演化的客观原因，也有商业营销放大概念的主观驱动，导致行业对世界模型的核心本质、技术边界与能力边界缺乏共识，资源投入分散在不同技术层级，核心方向不够聚焦，甚至出现“重表层效果、轻核心认知”的发展偏差。

1.3.2 重工程演示，轻底层机理

多数技术团队延续大模型时代的工程化思路，习惯于通过堆数据、堆算力、调参优化的方式推进技术迭代，聚焦于做演示、刷指标、优化视觉效果，而对世界模型的底层机理的研究投入不足。这种路径依赖本质上是“工程试错”范式的惯性延续，缺乏从第一性原理出发的理论创新。世界模型的核心突破点恰恰在于对认知机制的底层机理的重构，若始终停留在工程化优化层面，将难以突破大模型范式的能力天花板，最终陷入“局部优化、整体瓶颈”的发展困境。

1.3.3 学科壁垒深厚，产业生态割裂

世界模型是典型的多学科交叉领域，涉及认知神经科学、人工智能、数学、机器人学、芯片设计、控制理论等多个学科。但当前各学科领域相对独立，学科壁垒深厚：脑科学研究成果向工程算法转化不足，算法研发与芯片设计脱节，机器人本体与上层认知模型适配成本高，产学研用各环节之间缺乏有效衔接。软硬件生态割裂、跨学科人才匮乏、产业链协同不足，导致围绕世界模型的技术创新难以形成合力，制约了产业体系化发展。

1.3.4 价值认知偏差，发展预期失衡

当前业界对世界模型的价值认知呈现两极分化：一部分观点过度神化短期能力，高估技术落地速度，催生概念炒作与泡沫化风险；另一部分观点则将其视为远期科幻概念，低估技术演进速度与产业价值，导致战略投入不足。两种极端认知都不利于技术的理性发展。

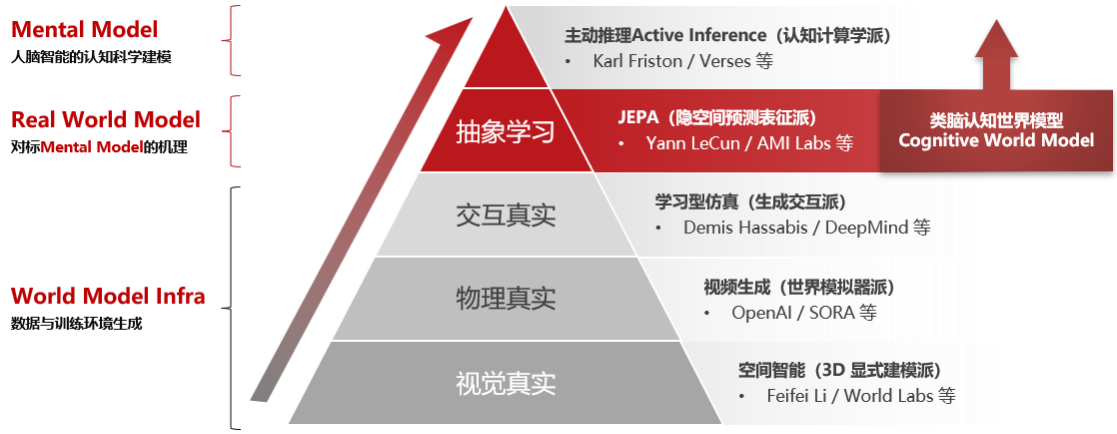
世界模型既不是短期快速落地的万能技术，也不是遥不可及的未来概念，而是需要以战略视角长期深耕、逐步突破的下一代 AI 核心基础设施，亟需形成客观理性的产业共识。

1.4 世界模型技术路线全景对比分析

当前，全球世界模型技术正处于多路线并行探索、分层级演进突破的关键阶段。形成了五大主流技术流派：空间智能派（3D 显式建模）、视频生成派（世界模拟器）、学习型仿真派（生成交互）、隐空间[1]预测表征派（JEPA 架构）、主动推理派（神经科学）。不同技术背景的机构从各自优势出发，形成了差异化的技术路径与流派分支。

整体来看，全球技术版图呈现“底层感知工程化提速、上层认知理论化攻坚”的分层发展格局：空间智能、视频生成等感知层技术依托大模型工程积累快速迭代，成熟度与商业化进度领先；抽象表征、主动推理等认知层技术深耕底层机理，长期成长空间更大，代表技术演进的核心方向。

图 1 世界模型技术流派解析



资料来源：具脑磐石，公开资料，编委会调研整理

1.4.1 五大主流技术流派深度解析

当前全球范围内，世界模型已形成五大主流技术流派，分别对应不同技术层级与核心目标，共同构成了世界模型的完整技术生态。

[1] 隐空间（又称潜在空间、潜空间）是深度学习与表征学习中的核心概念，指高维原始数据（图像、视频、文本、物理信号等）经过编码器网络非线性映射后，形成的低维、连续的抽象特征向量空间。
空间中的每一个向量（隐向量）对应原始数据的一个压缩表征，其维度不对应像素、字符等表层感官信息，而是编码了数据背后的核心语义、结构属性与内在规律；原始数据可通过解码器从隐空间近似重建。

1. 空间智能派：3D 显式建模路线

空间智能派是最早启动工程化探索的技术路线之一，核心是从视觉数据中重建环境的三维几何结构，构建显式化的空间世界表征。

- ✧ **技术内核**：以计算机三维视觉、多视图立体匹配、神经辐射场（NeRF）、3D 高斯溅射等技术为核心，通过单目、多目或视频输入，重建环境的几何结构、材质纹理与空间拓扑关系，实现高精度的三维空间表征。
- ✧ **代表力量**：李飞飞创立的 World Labs 为标杆机构，聚焦通用 3D 世界模型研发；英伟达 Omniverse 依托图形技术积累，构建工业级三维仿真与数字孪生底座；国内视觉团队也在消费级三维重建、场景建模领域快速跟进。
- ✧ **核心进展**：早期该路线以静态场景三维重建为核心能力，聚焦单一场景的几何结构还原；随着技术迭代，已逐步突破静态重建的边界，向动态时空模拟、可控视角生成与机器人仿真训练延伸。
- ✧ **产业定位**：世界模型的几何空间底座，为空间导航、物体操控、数字孪生提供精确的空间坐标系与几何基础，是实现“虚拟-真实”映射的关键支撑。
- ✧ **优势与短板**：核心优势是空间表征精度高、视觉真实感强，可直接支撑定位、抓取、路径规划等空间任务；核心短板是单独依靠空间几何表征难以获得物理因果推理、任务规划与主动学习能力，且大规模场景建模成本高、周期长，难以覆盖开放无限环境。

2. 视频生成派：世界模拟器路线

视频生成派是当前产业热度最高、公众认知最广的技术路线，核心是通过生成式 AI 学习视频数据的时空分布规律，生成符合物理视觉规律的连续画面，被业界寄予“世界模拟器”的期待。

- ✧ **技术内核**：以时空 Transformer、扩散模型为核心架构，在海量视频数据上进行大规模预训练，部分模型在像素空间工作，部分在压缩潜空间工作，在大规模视频上学习时空结构与条件生成，通过文本、图像等条件输入生成连续视频内容，本质是高维时空数据的统计拟合。
- ✧ **代表力量**：早期标杆为 OpenAI 的 Sora，曾首次确立“视频生成模型即世界模拟器”的产业认知。受公司战略向机器人领域倾斜等因素影响，Sora 独立应用与 API 已于 2026 年全面关停，该路线全球产业格局迎来重构。当前全球核心标杆为 Google DeepMind Veo 系列，已形成 Quality、Fast、Lite 三大分层版本，原生

支持音画同步，可精准模拟光影、刚体运动、流体等物理规律，最高支持 4K 分辨率，深度融入 Google 专业制片、消费生态与企业开发体系，走产业级落地路线。国内代表为字节跳动 Seedance 系列，2.0 版本跻身全球第一梯队，具备电影级画面质感与长镜头叙事连贯性，已落地影视制作、品牌营销等商用场景；目前正研发实时交互式视频生成技术，向可交互世界模拟器方向演进。

- ✧ **核心进展**：视频时长从秒级提升至分钟级，分辨率、流畅度与物理合理性持续优化，部分场景已能呈现基础的物体碰撞、重力运动等物理现象。
- ✧ **产业定位**：世界模型的视觉数据引擎，可提供大规模、多样化的视觉训练数据，也可作为数字孪生、内容创作的视觉渲染组件。
- ✧ **优势与短板**：核心优势是视觉保真度高、场景覆盖广、生成效率高，在内容创作、数字人驱动领域商业化落地快；核心短板是本质为像素级统计拟合，缺乏对物理规律的深度理解与因果推理，长时序生成易出现物理不一致、逻辑断裂问题，且无法支持精确的交互控制与行动反馈，不等同于可交互的具身世界模型。

3. 学习型仿真派：生成交互路线

学习型仿真派最贴近具身智能的交互本质，核心是让智能体在“行动-反馈”闭环中学习可用于预测与规划的环境动力学模型，并据此优化行为策略。

- ✧ **技术内核**：基于模型的强化学习（MBRL）通常从交互数据学习动作条件的动力学/奖励或价值模型，并在学习到的模型中进行规划或“想象”训练；它不必先构建高保真的外部虚拟仿真环境。Dreamer 等方法的关键正是学习紧凑的潜在动力学模型。
- ✧ **代表力量：**Google DeepMind 为技术引领者，Dreamer 系列模型持续迭代，大幅提升了基于模型的强化学习样本效率；OpenAI 机器人团队、波士顿动力等企业也依托仿真环境开展机器人控制策略学习。
- ✧ **核心进展**：MBRL 在若干控制基准上可显著提高交互效率，并已扩展到视觉控制、操作和导航等任务，已能在仿真环境中学习复杂操作、运动与导航策略，初步实现简单任务的 sim2real 迁移。
- ✧ **产业定位**：世界模型的**策略学习试验场**，是连接世界表征与行动控制的关键路径，为具身智能体提供低成本、高安全性的训练环境。
- ✧ **优势与短板**：核心优势是贴合具身智能的交互本质，支持端到端策略学习，可验证复杂任务；核心短板是虚拟环境与真实环境存在“虚实鸿沟”，物理特性、感官反馈差异大，迁移成本高；奖励函数设计依赖人工，探索效率与跨形态迁移能力

仍有较大瓶颈。

4. 隐空间预测表征派：JEPA 隐空间预测路线

隐空间预测表征派是世界模型从“像素级模拟”走向“语义级认知”的关键转折点，核心是在抽象隐空间中学习世界表征与预测规律，而非直接拟合原始输入。

- ✧ **技术内核**：以联合嵌入预测架构（JEPA）为核心，将多模态输入编码到抽象隐空间，通过预测缺失部分、未来帧的隐空间表征实现自监督学习。学习目标从“像素级重建”转向“语义级预测”，更贴合人脑的抽象认知模式。
- ✧ **代表力量**：图灵奖得主 Yann LeCun 创立的 AMI Labs 为核心引领者，将 JEPA 视为通往通用人工智能的核心架构；Meta AI 持续推进 I-JEPA、V-JEPA 等系列研究，验证了隐空间预测的效率优势。国内以具脑磐石为主要代表，基于 JEPA 架构融入类脑算法机制做多层次优化，打造基于类脑的认知世界模型。
- ✧ **核心进展**：已在图像、视频领域验证了 JEPA 在样本效率、泛化能力上显著优于逐像素重建方法，逐步向多模态融合、机器人控制场景延伸。
- ✧ **产业定位**：世界模型的**认知表征核心**，是实现高泛化、小样本学习的关键技术路径，构成了高层认知推理的表征基础。
- ✧ **优势与短板**：核心优势是样本效率高、泛化能力强、计算成本低，抽象表征更适合语义推理与任务规划；核心短板是工程化成熟度相对较低，向决策、行动控制层面的延伸仍在探索阶段，完整的端到端系统能力尚未验证。

5. 主动推理派：神经科学路线

- ✧ 主动推理派是当前类脑认知世界模型中理论体系最完整的路线之一。其核心不是复刻单一脑区，而是以概率生成模型和自由能框架统一感知推断、行动选择、主动探索与模型学习**技术内核**：以自由能原理（FEP）为底层公理，基于生成模型与变分贝叶斯推断，将感知更新、行动选择、模型学习统一到“最小化自由能”的同一目标下。智能体通过最小化变分自由能更新认知，通过最小化期望自由能选择行动，形成完整的自闭环。
- ✧ **代表力量**：Karl Friston（英国皇家学会院士，FRS；同时为 Academia Europaea 成员）担任 VERSES 首席科学家，推动主动推理的工程化探索；国内以具脑磐石为代表的企业，深度融合主动推理与 JEPA 架构，打造面向具身智能的类脑认知世界模型；全球顶尖高校与脑科学研究机构是基础理论的核心供给方。

- ✧ **核心进展**：已形成离散推理、连续推理、混合推理等成熟建模框架，在机器人控制、自适应系统等场景完成初步验证，理论体系持续完善。
- ✧ **产业定位**：世界模型的**顶层认知框架**，为构建完整的具身认知系统提供统一的理论范式与计算体系，代表世界模型的长期演进方向。
- ✧ **优势与短板**：核心优势是理论框架统一，可显式表示隐状态、不确定性、偏好与信息价值，具有结构化建模、主动探索和可解释推理的潜力；核心短板是大规模、高维连续控制的可扩展性仍待突破，工程化工具链和人才储备也相对不足。

1.4.2 各技术流派深度对比分析

五大流派并非平行并列关系，而是处于不同技术层级，分别对应世界模型从感知到认知的不同能力模块。

图表 1 全球五大世界模型技术流派多维度对比

对比维度	空间智能派	视频生成派	学习型仿真派	抽象表征派	主动推理派
技术层级	感知层-空间表征	感知层-视觉模拟	交互层-策略学习	认知层-表征学习	认知层-全闭环架构
核心方法论	三维重建与显式几何建模	扩散模型与时序统计拟合	基于模型的强化学习	隐空间联合嵌入预测	自由能原理与主动推理
核心目标	构建精确 3D 空间环境	生成逼真视觉内容	学习交互行为策略	学习抽象世界表征	统一感知行动学习闭环
代表机构	World Labs、英伟达	OpenAI、国内大模型厂商	DeepMind、OpenAI	AMI Labs、Meta、具脑磐石	Verses、具脑磐石
数据依赖度	中高，依赖多视角/视频数据	极高，依赖海量视频数据	高，依赖仿真交互数据	中，自监督降低标注依赖	较低，强调在线交互与先验结构
因果推理能力	单独几何模型通常不提供因果识别	低，主要学习观测时空规律	中，支持动作条件动力学与干预	中低，以表征预测为主	中高，可在生成模型中显式纳入因果结构与策略干预
泛化能力	中等，空间泛化强	较弱，分布外易失真	中等，特定场景强	较强，语义泛化好	强，跨场景通用泛化
终身学习能力	弱，重建后固定	无，训练后静态	较弱，环境依赖强	中等，支持增量表征	强，可支持在线更新，稳定性取决于

					记忆与学习机制
可解释性	高，几何结构显式可查	差，黑箱生成过程	较差，策略可解释性弱	中等，表征可分析	高，结构层具有较高可解释性，底层神经模块仍可能较复杂
工程成熟度	较高	中高	中等	中低	较低
商业化阶段	场景落地初期	内容商业化快速起量	科研向产业过渡	前沿技术验证	早期探索阶段
核心价值定位	世界模型的几何底座	世界模型的视觉引擎	世界模型的策略训练场	世界模型的表征核心	世界模型的认知框架

资料来源：公开资料，编委会调研整理

从系统视角看，完整具身世界模型往往同时需要几何与对象表征、动作条件动力学、记忆、不确定性估计以及规划控制。本文五类路线可视为互补的技术模块，在不同系统中承担的功能层级可有所不同。

从发展路径看，底层感知流派工程成熟度更高、商业化更早，但能力天花板明显；高层认知流派技术难度更大、成熟度更低，但代表长期发展方向。

1.4.3 技术演进趋势：分层成熟与跨流派融合

全球世界模型技术演进呈现两大清晰趋势：底层感知技术加速工程化成熟，上层认知技术逐步向类脑认知方向收敛；同时跨流派深度融合，形成“认知为核、感知为用”的整体演进方向。

1. 分层演进：底层快速落地，上层攻坚突破

- ✧ **感知层技术**：空间智能、视频生成受益于大模型技术外溢与工程化积累，迭代速度快、商业化进程领先，将率先在内容创作、数字孪生、视觉仿真等领域实现规模化落地，持续优化视觉效果与生成效率。
- ✧ **认知层技术**：抽象表征、主动推理处于技术突破前期，研发周期长、难度大，但代表了下一代 AI 的演进方向。随着脑科学研究深入与工程能力提升，认知层技术将逐步迎来突破，从根本上提升世界模型的泛化、推理与学习能力。

2. 跨流派融合：多技术互补协同

单一技术路线难以支撑完整的世界模型能力，未来必然走向多技术深度融合，核心融合路径包括：

- ✧ **空间智能+抽象表征**：将 3D 显式建模的几何精度与 JEPA 的抽象泛化能力结合，构建兼具空间精度与语义泛化的空间认知模型，提升机器人导航与场景理解能力。
- ✧ **视频生成+因果建模**：在视频生成模型中融入物理因果约束，提升长时序生成的物理一致性与逻辑连贯性，从“视觉生成器”向“物理模拟器”升级。
- ✧ **学习型仿真+主动推理**：通过显式建模先验偏好、信息增益和不确定性，可将目标实现与主动探索纳入统一策略选择过程。主动推理通常以先验偏好或学习到的偏好表示目标，可与 MBRL、MPC 和 RL 形成互补或混合架构。
- ✧ **全栈融合架构**：以类脑认知机制为顶层核心，向下整合空间表征、视觉生成、仿真交互等能力，构建“感知-认知-决策-行动”全链路完整世界模型体系。

3. 长期走向：类脑认知成为核心收敛方向

从技术演进底层逻辑看，脑启发机制在多类技术路线中的认可度持续提高：Yann LeCun 的 JEPA 借鉴人脑抽象预测机制，DeepMind 在模型中引入记忆系统与预测编码，越来越多的团队从脑科学中寻找灵感。

长期来看，世界模型的技术突破将从“工程驱动”转向“机理驱动”，类脑认知机制将成为核心创新源泉。实现脑科学机理与工程实现的深度融合，将会是下一代世界模型的核心方向。

第二章 理论溯源：人脑如何构建自身的“认知世界模型”

2.1 演进脉络：从哲学猜想到计算科学体系

回顾人工智能发展史，人类对“世界模型”的认知经历了从哲学猜想、心理学假说、神经科学机制到计算科学框架的逐步深化过程，形成了完整的思想传承链条。

1. 思想萌芽阶段：无意识推理与主动知觉（19 世纪）

19 世纪德国物理学家、生理学家 Hermann von Helmholtz 提出“无意识推理（unconscious inference）”思想，强调知觉会利用既有经验和隐含推断解释不完整的感觉输入。Helmholtz 也对实验心理学产生深远影响，但将其主要身份直接写为“心理学家”不够准确。该思想可视为现代生成式知觉与贝叶斯脑观念的重要历史先声之一。

2. 概念成形阶段：心智模型的正式提出（20 世纪中期）

1943 年，英国心理学家肯尼思·克雷克在《解释的本质》一书中正式提出“心智模型”（Mental Model）概念。他认为，人类大脑会在内部构建外部现实的“小型缩微模型”，该模型不是对物理世界的精确复制，而是对环境关键要素、因果关系与动态规律的抽象表征。大脑通过该模型预测事件发展、进行逻辑推理、做出行动决策。心智模型的提出，首次明确了心智模型的核心功能与存在价值，为后世的世界模型奠定了概念基础。

3. 机制验证阶段：预测编码的神经实证（20 世纪后期）

20 世纪后期，预测编码逐步发展为现代计算框架。Srinivasan、Laughlin 与 Dubs 于 1982 年从视网膜冗余抑制角度提出“predictive coding”；Mumford 等随后发展层级皮层计算观点；Rao 与 Ballard 于 1999 年提出高层反馈携带预测、前馈通路携带残差信号的视觉皮层模型，成为现代层级预测编码的重要里程碑。预测编码由此形成一组具有实验支持但实现形式多样的理论模型。

预测编码为解释皮层中的自上而下预测与自下而上误差信号提供了重要计算假说，并得到多类实验现象的支持；同时，皮层感知还涉及前馈编码、注意、递归处理等多种机制，预测编码并非唯一解释框架。

4. 体系统一阶段：主动推理理论（21 世纪以来）

进入 21 世纪，Karl Friston 等发展了自由能原理与主动推理框架，将贝叶斯推断、生成模型、感知与行动纳入统一的规范性形式，为理解认知与具身行为提供可计算理论，也

成为类脑智能的重要理论来源。

2.2 神经机制基础：多脑区协同的认知系统

人脑所具备的“认知世界模型”并非单一脑区的功能，而是新皮层、海马系统、小脑与多个功能网络协同工作的结果。不同脑区承担不同功能层级，通过双向连接形成动态闭环，共同实现对世界的表征、预测与推演。

1. 新皮层：层级化预测表征核心

新皮层在哺乳动物高级感觉、联合与认知处理中具有重要作用，典型新皮层具有六层组织并呈现丰富的层级与回返连接。其分层表征和回返交互为世界模型的层级状态表征提供了重要神经计算启发。

- ✧ **层级表征体系：**从初级感觉皮层（视觉、听觉、体感）到次级联合皮层，再到高级前额叶联合皮层，信息处理逐层抽象：低层编码具体感官特征，中层编码物体与事件，高层编码抽象概念、因果规则与任务目标。这种层级结构天然支持从具体到抽象的多尺度世界表征。
- ✧ **预测编码运行：**预测编码假说认为，皮层层级可能通过自上而下预测与自下而上误差信号交互更新内部表征。不同脑区是否普遍采用同一实现，以及具体细胞/层级对应关系仍未完全确定
- ✧ **高级认知功能：**高级认知功能由前额叶、顶叶、内侧颞叶、基底神经节等分布式网络共同支持。前额叶及其与顶叶、海马和基底神经节构成的分布式网络，是目标维持、规划与认知控制的重要神经基础，可为世界模型中的高层状态维持、任务规划和反事实推演提供机制启发。

2. 海马-内侧颞叶系统：时空记忆与认知地图载体

海马体与内侧颞叶周边脑区共同构成记忆与空间认知系统，是世界模型的“记忆库”与“空间坐标系”。

- ✧ **空间认知地图：**海马体中的位置细胞、内嗅皮层中的网格细胞，能够编码环境中的位置、方向与拓扑关系，在大脑中构建环境的“认知地图”，支撑智能体的空间导航与场景定位。这一机制正是托尔曼认知地图理论[2]的神经基础。

[2] 托尔曼认知地图理论是认知心理学的奠基性理论，也是“世界模型”思想在心理学领域的最早理论雏形，由美国心理学家爱德华·托尔曼（Edward C. Tolman）于 1948 年在《Cognitive maps in rats and men》中正式提出，核心是反驳当时主流的行为主义“刺激 - 反应”学习观，首次系统论证了智能体通过

- ✧ **情景与时序记忆：**海马体负责编码事件的时间、地点、人物与情节，形成连贯的情景记忆，保存世界交互的历史经验。
- ✧ **记忆巩固机制：**记忆巩固涉及海马-新皮层反复交互，尖波-涟漪及清醒/睡眠重放在其中发挥重要作用。其“快速情景记忆+慢速皮层学习”的双系统特征，为人工系统的持续学习提供直接启发。

3. 小脑：运动与认知的前馈预测器

小脑在运动时序、误差校正和内部前向模型方面具有较充分证据，也与部分认知任务相关。其快速前向预测机制可为具身世界模型中的身体动力学预测和误差校正提供工程启发。

- ✧ **内部前馈模型：**小脑构建身体与环境的动力学模型，能够根据当前状态与动作指令，提前预测下一刻的身体姿态与感官反馈，在误差出现前就修正动作，大幅提升运动控制的速度与精度。
- ✧ **认知预测功能：**小脑还参与时序预测、概率推理与认知模拟，支持对物理过程的快速推演，是世界模型动力学预测能力的重要组成部分。

4. 多网络协同工作机制

除核心脑区外，额顶控制网络、默认模式网络、基底神经节等功能网络协同参与：额顶网络负责任务执行中的认知控制与注意力调配；默认模式网络支持静息状态下的记忆回放、反事实想象与思维模拟；基底神经节负责习惯形成与策略选择。多脑区、多网络的动态协同，共同构成了人脑世界模型的完整神经基础。

构建环境的内部抽象表征指导行为的核心逻辑。

2.3 核心能力维度：世界模型的五大认知支柱

基于多脑区协同机制，人脑“认知世界模型”形成了五大核心能力，共同支撑人类灵活高效的具身智能与通用认知能力。

1. 层级化感知抽象能力

人脑能够从海量高维的多模态感官输入中，快速提取关键特征，形成低维抽象表征，用极小的信息代价编码高维世界。

- ✧ **底层机制：**分层级特征编码、稀疏神经元激活、注意力门控选择。大脑并非均等处理所有输入，而是根据任务目标选择性提取关键信息，忽略无关细节。
- ✧ **核心价值：**大幅降低信息处理复杂度，实现跨模态统一表征，同时提升模型的泛化能力，支持从有限经验中提炼通用规律。

2. 结构化时空记忆能力

人脑“认知世界模型”具备完整的时间与空间维度，能够存储、组织与检索历史交互经验，形成结构化的记忆体系。

- ✧ **空间维度：**以拓扑认知地图的形式编码环境结构，支持空间推理与导航；
- ✧ **时间维度：**以事件序列的形式编码交互过程，支持时序推演与因果追溯；
- ✧ **核心特点：**记忆并非静态存储，而是动态重构过程，大脑可根据线索重组记忆片段，兼顾稳定性与灵活性，支持知识迁移与场景适配。

3. 直觉化物理因果推演能力

人脑“认知世界模型”天然具备对物理世界规律的直观理解，能够不依赖复杂计算，快速推演物体运动、碰撞、受力变化等物理过程，预判事件发展走向。

- ✧ **底层机制：**大脑通过长期交互经验，内化形成了“直觉物理引擎”，能够在内部快速模拟物理过程，输出定性的因果判断与定量的趋势预测。
- ✧ **核心价值：**支撑对行动后果的预判、对环境变化的响应，是决策与规划的认知基础，也是人类能够灵活应对开放物理世界的关键。

4. 生成式反事实想象能力

人脑“认知世界模型”不仅能够表征真实世界，还能够建构虚拟场景，进行反事实推理与思维实验。

- ✧ **底层机制：**默认模式网络、海马及前额叶等网络共同参与情景构建、未来想象、自传体记忆和反事实推演，体现出跨网络协同的分布式机制。
- ✧ **核心价值**：在内部模型中低成本完成多方案试错与比较，支持创造性问题解决、备

选方案规划与风险预判，是人类智能超越“即时反应”的核心标志。

5. 闭环式主动探索学习能力

人脑“认知世界模型”是面向行动的开放系统，通过“预测-行动-感知-更新”的持续闭环，实现自主探索与终身学习。

- ✧ **底层机制：**行动不仅服务于目标达成，也服务于信息获取。大脑会主动选择不确定性高的区域进行探索，通过新的观测修正预测误差，更新内部模型。
- ✧ **核心特点：**自动平衡“利用”与“探索”，在稳定完成任务的同时持续拓展认知边界，实现能力的终身持续进化。

2.4 统一理论框架：主动推理与自由能原理

主动推理（Active Inference）是当前最具统一性的规范性计算框架之一，以概率生成模型、变分推断和自由能相关目标统一表述感知、学习与行动，可为类脑认知世界模型提供重要的理论与算法工具。

2.4.1 自由能原理：生命智能的底层公理

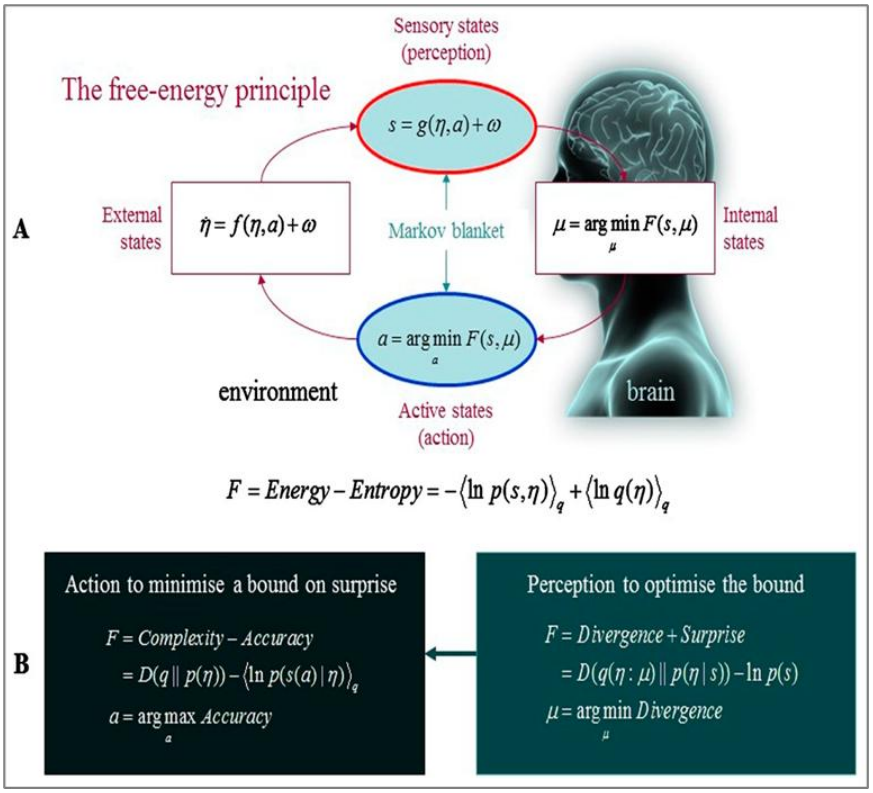
自由能原理（Free Energy Principle, FEP）提出，在特定统计与动力学假设下，可将维持有限状态集合的自组织系统表述为最小化变分自由能。自由能原理的关键不是简单降低“预测误差”，而是在生成模型下同时约束解释观测的准确性与偏离先验的复杂性。由此，感知可以被表述为对隐藏世界状态的推断，而行动则进一步改变未来观测，使智能体形成“预测—感知—行动—更新”的闭环。

在变分推断中，变分自由能是负对数模型证据（surprisal，常译“惊异”）的上界；surprisal 定义为$-\log p(o)$，并不等同于简单的“观测与预测的偏差”。在常用分解下，自由能可写为“复杂性−准确性”，其中各项均相对于特定生成模型和近似后验定义：

$F[q] = D_{KL}[q(s) \| p(s)] - E_q[\log p(o|s)]$。其中，“复杂性”对应近似后验 $q(s)$相对先验 $p(s)$的 KL 散度项，“准确性”对应后验下的期望对数似然项。这里的复杂性不是网络参数量，准确性也不是普通分类任务中的准确率。

Karl Friston 在《**The free-energy principle: a unified brain theory?** 》的 Nature 文章里有清晰的论述。

图 2 自由能原理（Free Energy Principle, FEP）图示



资料来源：Karl Friston，《The free-energy principle: a unified brain theory? 》，编委会整理

最小化自由能，本质是在保证模型预测精度的同时，尽可能降低模型复杂度，实现效率与效果的最优平衡。这一原理揭示了生命智能的底层运行逻辑：所有认知与行为，最终都服务于减少世界的不确定性，让外部环境符合内部预期。由此，智能体既可通过更新内部信念解释当前观测，也可通过行动改变未来观测，使内部模型与感觉证据持续协调。

2.4.2 主动推理理论的核心内涵与技术价值

1. 主动推理理论的核心内涵

主动推理（Active Inference）是基于概率生成模型和变分推断的规范性计算框架。在 Karl Friston 等的最小自由能原则的理论框架里，当前观测下的隐状态推断可通过最小化变分自由能实现；策略选择则通过最小化期望自由能（或等价形式）综合偏好满足与认识性价值，因此同时涉及目标实现与信息获取。由此基于最小自由能原理、预测编码和贝叶斯理论的基础上发展出的理论假说。用于解释和建模生命有机体如何在不确定环境中通过感知、行动和学习维持可生存状态也取得重要进展。这个假说可表述为：生命智能体通过最小化变分自由能来近似后验推断，通过最小化期望自由能来选择行动，从而令感知观察的

“惊异（Surprise）”最小化。

2. 主动推理的双视角逻辑

要系统理解主动推理理论，可以从自上而下和自下而上两个逻辑角度去审视（如下图）。

图 3 主动推理（Active Inference）自上而下和自下而上的两套逻辑视角



资料来源：《Active inference: The Free Energy Principle in mind, Brain, and Behavior》，编委会整理

从顶层理论来看，就解释如人类这种高等智能体本源而言，主动推理（Active Inference）理论从“自由能最小化”出发能引申出一系列形式及实证假设，可用于解释大量认知科学与神经科学问题。包括神经集群如何编码自由能最小化涉及的变量，最小化自由能的过程如何对应于特定认知过程，如知觉、行动选择和学习，以及主动推理的主体为推进自由能的最小化会采取哪些行动等。主动推理架构之“主动”在于智能体通过行动主动地观察采样世界。智能体必须对其“行动-知觉”环路施以适应性的控制，由此获得带认知偏好的感知观察。

因此，在计算神经科学领域，该理论也是**指导世界模型算法创新的重要理论源泉**。

从底层计算来看，主动推理架构的底层逻辑从贝叶斯大脑的理念出发，该理念视大脑

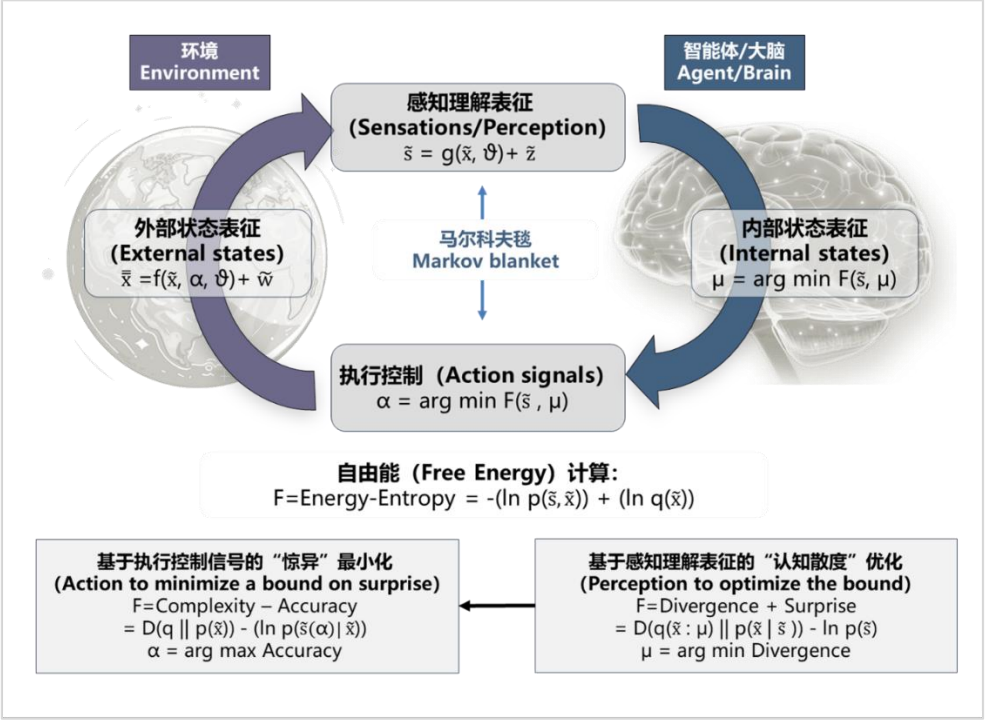
为一台推理引擎，致力于优化感知刺激之诱因的概率表征。它把知觉、行动选择和学习统一到生成模型、变分自由能和期望自由能的计算之中：知觉更新隐状态信念，学习更新模型参数，行动选择兼顾偏好实现、风险降低和信息增益。这一框架对世界模型的价值在于，它天然把不确定性、探索、身体状态和目标偏好纳入同一个闭环。主动推理是（原本不可解的）推理问题的一种特殊的变分近似，这种近似有生物学意义上的可行性。主动推理的底层逻辑有助于阐释主体如何最小化自由能——这意味着主动推理不仅是一条原则，还是对认知功能及其神经基础的机制解释。从计算角度看，主动推理以贝叶斯生成模型为核心：知觉更新隐状态信念，学习更新模型参数，行动选择同时考虑偏好实现、风险或歧义以及信息增益。其关键价值在于把不确定性、探索、身体状态和目标偏好纳入同一闭环，使“感知什么、相信什么、采取什么行动”成为同一推断问题的不同侧面。

3. 主动推理的计算体系与技术价值

主动推理（Active Inference）理论既是一套解释生命体或具身智能体感知-行动闭环的规范框架，也提供了基于离散推理、连续推理及混合推理模型的计算建模框架。

主动推理理论以**自由能原则（FEP）为底层理论原则，用贝叶斯大脑模型**诠释世界模型的计算本质，融合生成模型与**预测编码**给出计算神经建模的算法实现，统一了世界模型的感知、行动与学习闭环，为世界模型提供“预测-误差-行动-更新”的通用算法框架（如下图）。

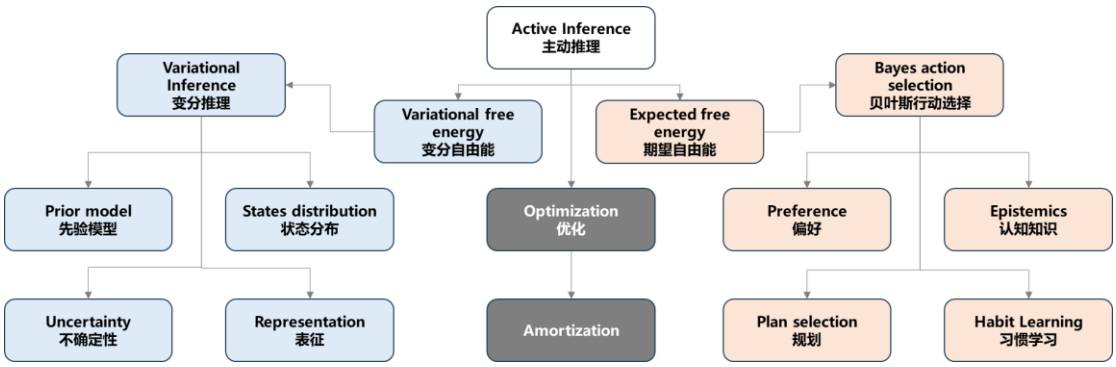
图 4 智能体基于自由能原则（FEP）的主动推理



资料来源：《Active inference: The Free Energy Principle in mind, Brain, and Behavior》，编委会整理

主动推理认知计算框架下的自由能原则（FEP）、预测编码理论、贝叶斯大脑模型等核心类脑认知理论及其基于**变分自由能**与**期望自由能**的认知推理计算体系（如下图），共同为**认知世界模型（Cognitive World Model，Cog-WM）**提供了坚实的科学底层理论支撑，让 AI 算法范式创新从“工程试错”转向为“脑科学理论驱动”。把脑科学中的可计算约束转化为算法目标函数、网络结构、学习规则和能耗指标，使具身智能从单纯工程试错转向“理论约束下的工程搜索”。 FEP、预测编码、贝叶斯脑与主动推理可为 Cog-WM 提供不同层次的可计算约束。工程实现可将其中可检验的机制转化为目标函数、网络结构、学习规则或能效约束，并通过统一任务上的消融和对照实验评估各机制的独立贡献。

图 5 基于变分自由能与期望自由能的主动推理计算体系



资料来源： 《Active inference: The Free Energy Principle in mind, Brain, and Behavior》，编委会整理

主动推理提供了离散推理、连续推理、混合推理三种可工程化的建模框架，核心围绕变分自由能与期望自由能两大计算量展开：

- ✧ **变分自由能：**基于当前观测与内部信念计算，驱动知觉更新与模型学习；
- ✧ **期望自由能：**基于未来策略与预期观测计算，驱动行动选择与规划。

对于具身智能而言，主动推理框架的核心价值在于：它天然将身体状态、环境不确定性、能量约束、任务偏好纳入统一计算，无需人工设计奖励函数，能够支撑智能体在开放动态环境中的自主运行与终身学习，是最接近人脑运行机制的认知计算框架之一。对于具身智能，主动推理的价值还在于能够在生成模型中显式表示隐状态、不确定性、偏好和信息价值。身体状态和能量约束可作为生成模型中的显式变量或约束纳入计算；主动推理不依赖传统强化学习中单一的显式标量奖励函数，而是通过先验偏好或学习得到的目标分布表达价值，并以期望自由能统一权衡目标实现与信息获取。

2.5 脑科学对世界模型的启示

立足当下，回顾历史和展望未来，我们都能清晰看到人脑的结构和功能神经机制为世界模型的技术突破提供了第一性原理指引，也指明了跳出当前大模型范式瓶颈的核心方向。

1. 本质转向：从“模拟感官世界”到“模拟认知系统”

人脑的核心是抽象认知推演，而非像素级感官复刻。这启示我们：世界模型不能停留在视频生成、3D 渲染等表层视觉模拟，必须聚焦因果表征、时序记忆、决策规划等核心认知能力，打造面向行动的内部认知推演系统。

2. 机制转向：从“批量数据拟合”到“预测误差闭环”

人脑通过“预测-误差-更新”的交互闭环实现学习，而非一次性批量数据拟合。这启示我们：世界模型应从“离线预训练范式”转向“预测编码范式”，在与环境的持续交互中通过预测误差驱动模型更新，从根本上降低“数据依赖”和“灾难性遗忘”。脑学习包含在线适应、离线巩固、监督或强化信号以及多时间尺度可塑性。对人工世界模型的启示是把离线预训练与在线预测误差驱动更新结合起来，并显式处理稳定性—可塑性权衡，以提高持续学习中的知识保持与新知识吸收能力。

3. 架构转向：从“扁平大模型”到“层级化抽象表征”

人脑采用分层级信息处理架构，实现从具体感知到抽象认知的递进。这启示我们：世界模型应构建分层级表征体系，在隐空间实现多模态融合与抽象，而非在像素、token 层面做稠密计算，以此提升跨场景泛化能力与计算效率。

4. 能力转向：从“静态训练模型”到“内置终身记忆”

人脑依靠海马-新皮层双系统与离线巩固机制实现持续学习。这启示我们：世界模型必须内置专门的记忆系统与巩固机制，突破“预训练+微调”的静态模式，支持增量学习、知识迁移与持续进化，克服灾难性遗忘问题。

5. 效率转向：从“稠密暴力计算”到“稀疏能效优化”

人脑在有限能量预算下，依靠稀疏编码、事件驱动实现高效智能。这启示我们：世界模型应引入类脑稀疏计算与能量约束机制，打破稠密计算范式，大幅降低功耗与时延，适配端侧具身智能的部署需求。

第三章 类脑视角下世界模型的定义、架构与能力边界

3.1 四类易混淆的“模型”概念

当前业界对世界模型的认知存在诸多混淆，本小节从核心目标、技术本质、能力边界三个维度，对四类常见的“世界模型”易混淆概念进行明确区分。

图表 2 市场上主要的几类“模型”概念对比分析

概念类型	核心目标	技术本质	核心优势	核心短板	产业定位
视频生成模型	生成逼真视频画面	像素级时空统计拟合	视觉保真度高	缺乏因果推理与交互能力	世界模型的数据/视觉组件
三维仿真引擎	构建精确虚拟环境	显式物理建模与图形渲染	空间与物理精度高	规则人工设定，自主学习弱	世界模型的仿真支撑工具
生成式世界模型	通用环境状态表征预测	多技术路线的数据驱动建模	场景适配性较广	因果弱、数据依赖高、可解释性差	世界模型的主流初级形态
类脑认知世界模型	构建类人认知推演系统	脑启发的认知计算架构	低数据、高泛化、终身学习、低功耗	技术复杂度高，工程落地难度大	世界模型的高级发展形态

资料来源：公开资料，编委会调研整理

- **视频生成模型**：以 OpenAI Sora、Runway Gen-2 等为代表，核心是生成逼真视频内容，擅长视觉创作与渲染，但不具备深度物理理解与因果推理能力，长时序生成易出现物理不一致问题，是世界模型的重要视觉组件而非完整世界模型。
- **三维仿真引擎**：以 Unity、Unreal Engine、Isaac Gym 等为代表，核心是构建精确 3D 虚拟环境，具备高精度空间表征与物理模拟能力，但物理规则需人工编写，无法自主学习规律，是世界模型的重要支撑工具。
- **生成式世界模型**：旨在构建适配多场景多任务的通用环境模型，技术路线多样，但多数延续大模型数据驱动范式，存在因果推理弱、可解释性差等问题，是当前行业主流初级形态。
- **类脑认知世界模型**：借鉴人脑认知机制，构建面向行动的内部认知推演系统，以提高

数据压缩效率、开放环境泛化、持续学习和运行能效为核心技术目标，聚焦抽象表征与推理能力，具备低数据依赖、高泛化、终身学习、低功耗等差异化优势，是世界模型的高级发展形态。

3.2 类脑认知世界模型的核心定义

结合**人脑智能本源**与类脑技术发展趋势，本白皮书对**类脑认知世界模型（Cognitive World Model, Cog-WM）**定义如下：

类脑认知世界模型是一套以人脑主动推理与预测编码理论为核心，系统借鉴人脑结构和认知神经机制，面向具身智能体在开放物理世界的交互、移动、操作、协作等全栈能力构建的认知推演系统。它在隐空间中联合表征外部环境感知状态、内部认知预测状态、任务时空记忆、行动后果与不确定性，通过“预测-误差-行动-更新”的闭环机制，实现感知、认知、因果推演、时空记忆与自主决策，支撑智能体在开放环境中的持续学习与自主交互。

相较于传统世界模型（含三维仿真引擎、视频生成模型/ 生成式世界模型等主流工程化路线）与 VLA 具身模型，类脑认知世界模型（Cognitive World Model, Cog-WM）代表的是具身智能 2.0 的技术方向，其包含五大核心内涵：分别是**理论根基、核心本质、表征方式、运行机制、能力目标**，具体展开如下：

（1）理论根基：以脑科学与认知神经科学机理为第一性原理，而非统计学习理论

类脑认知世界模型的底层理论根基并非来自深度学习与统计学习框架，而是以认知神经科学的成熟机理为第一性原理，核心依托自由能原理、大脑预测编码机制、海马体认知地图结构、新皮层层级处理逻辑等生物智能工作机理，从“大脑如何构建世界认知”的底层规律出发，反向推导人工智能世界模型的架构设计与算法逻辑，具备明确的生物可解释性与机理驱动性。

与之相对，传统大模型驱动的世界模型以统计学习理论、深度神经网络架构为理论根基，核心逻辑是通过大规模参数拟合训练数据的分布规律，以数据驱动为核心导向，不追求底层机理的可解释性，仅以输出效果与拟合精度为优化目标。

（2）核心本质：主动认知推演系统，而非数据拟合映射系统

类脑认知世界模型的核心本质是服务于智能体自主决策的内部认知推演系统，以最小化智能体与环境交互的自由能为核心目标，通过内化环境的因果逻辑、时空演化与交互约束，在内部完成状态预测、行动规划、反事实推理等认知过程，核心价值指向“支撑决策”，是智能体的“内部认知中枢”。

传统世界模型与 VLA 本质上是输入 - 输出型的数据拟合映射系统：生成式世界模型以

拟合视觉数据分布为核心，输出符合统计规律的图像、视频等感官内容；VLA 具身模型以拟合感知 - 动作映射关系为核心，输出匹配任务的动作指令。二者均以“生成 / 响应”为核心价值，不具备真正意义上的自主内部认知推演能力。

（3）表征方式：分层抽象的认知级表征，而非感官级数据表征

类脑认知世界模型采用对应大脑新皮层层级结构的分层抽象表征：底层编码初级感知特征，中层编码物体、空间等结构化信息，高层编码抽象的因果规律、语义概念与时空模式。其表征体系（对应隐空间）以“预测有效性”为优化目标，主动丢弃与认知推演无关的感官细节，仅保留对决策、预测有价值的核心结构信息，属于“认知级表征”。

传统世界模型同样具备表征体系，但属于面向还原精度的感官级数据表征：生成式模型的隐空间以“重建相似度”为目标，尽可能保留像素、纹理等表层感官信息；VLA 的表征以“动作拟合精度”为目标，编码感知与动作的统计关联。二者均聚焦数据层面的特征拟合，不追求对环境底层因果结构的抽象编码。

（4）运行机制：预测 - 误差 - 更新自主内闭环，而非外部驱动的开环 / 外闭环

类脑认知世界模型遵循主动推理框架，以“自上而下预测→自下而上误差→模型自主更新”的内部自闭环为核心运行机制。模型持续主动生成对环境状态与感官输入的预测，通过实际感知与预测的偏差自动修正内部表征与模型参数，整个闭环无需外部任务指令即可自主运行，本质是一个主动的“预期系统”， 这也为为终身学习提供了内生机制。

传统路线不具备这种自主内部闭环：生成式世界模型是“输入提示→生成输出”的单向开环过程，模型自身不会主动生成内容或自我更新；VLA 具身模型是“感知输入→模型推理→动作输出→外部反馈”的外部驱动闭环，运行高度依赖任务指令与环境奖励，模型参数不会自主迭代，本质是被动的“响应系统”。

（5）能力目标：终身自主的低功耗认知能力，而非固定范式的高精度拟合能力

类脑认知世界模型的最终能力目标，是实现低数据依赖、高泛化能力、持续终身学习、低功耗推理四大核心能力：以预测误差为天然学习信号，无需大规模标注数据与密集外部奖励，通过少量环境交互即可构建有效的环境认知表征，支持小样本快速适配全新场景；能基于抽象表征推演未见过的场景，具备强迁移能力；可在与环境的持续交互中自主学习新知识，无需大规模重训；仅在环境变化产生预测误差时触发计算，天然低功耗；最终目标是支撑智能体在开放、动态、未知环境中实现高实时性的自主决策与交互。

传统世界模型与 VLA 的能力目标聚焦于固定范式下的高精度拟合：追求训练分布内的生成质量、动作精度与任务准确率，但模型部署后参数基本固定，增量学习能力弱，适配新

场景依赖大量数据微调；采用全连接密集计算范式，算力开销高且固定；泛化能力严格受限于训练数据分布，对未知场景的迁移能力弱。

3.3 类脑认知世界模型五层技术架构

类脑认知世界模型采用分层级技术架构，对应人脑认知处理层级，从下到上依次为：多模态感知抽象层、物理规律与因果表征层、时空记忆与认知地图层、内部沙盘与反事实推演层、主动推理与行动规划闭环层。越底层偏向基础感知表征，越上层偏向高级认知决策，而双向连接、协同工作的机制，确保了信息既能自下而上地进行高维抽象，决策也能自上而下地进行预测引导，这也是实现通用人工智能（AGI）极具前景的技术路径。

3.3.1 多模态感知抽象层（基础感知）

作为模型的输入基石，它类似于人类的感官系统。负责将视觉、听觉、触觉等多源异构的物理信号，提炼、融合并抽象为机器可理解的高维语义特征。对应人脑初级感觉皮层与联合感觉皮层，负责处理多模态感官输入，提取关键特征，形成统一抽象感知表征。

- ✧ **核心功能：**多模态信息编码、跨模态对齐融合、主动感知选择。
- ✧ **技术关键：**借鉴人脑分层编码与稀疏表征机制，采用类脑感知编解码架构，实现高效特征提取与多模态融合，降低数据维度，提升感知泛化能力。

3.3.2 物理规律与因果表征层（世界常识）

在感知之上，模型开始理解“世界是如何运转的”。它不仅进行静态的模式识别，更致力于建模物理定律（如重力、碰撞、客体恒常性）和事件之间的因果链条，赋予 AI 类似于人类直觉物理（Intuitive Physics）的常识。对应人脑高级联合皮层，负责学习与表征物理世界运行规律、物体属性与因果关系，构建结构化世界知识。

- ✧ **核心功能：**物理规律表征、因果关系建模、物体属性与可供性表征。
- ✧ **技术关键：**基于生成模型与预测编码，从交互数据中自主学习物理规律与因果结构，而非人工预设规则，实现开放环境自适应。

3.3.3 时空记忆与认知地图层（时空定位）

引入中长期时空尺度。通过模拟大脑海马体与内嗅皮层的机制，构建环境的“认知地图”与情节记忆，使模型能够在复杂的时空背景下进行自我定位、导航与情境关联。人脑海马体与内侧颞叶系统在情景记忆、关系表征和空间导航中的功能，为这一层的时空组织机制提供重要神经启发。

- ✧ **核心功能：**空间认知地图构建、时序情景记忆存储、记忆巩固与检索。
- ✧ **技术关键：**借鉴海马体-新皮层互补记忆架构，构建分层级记忆系统，实现持续增量学习，克服灾难性遗忘。

3.3.4 内部沙盘与反事实推演层（想象与模拟）

这是认知能力向高级阶段跃升的关键。模型无需在真实物理世界中以高昂代价去“试错”，而是可以在内部“沙盘”中，针对未发生的情况进行“如果……会怎样”的反事实推演（Counterfactual Thinking），预测不同决策的未来走向。对应人脑前额叶皮层与默认模式网络，负责在内部模型中进行模拟推演、反事实想象与思维实验，评估不同行动后果。

- ✧ **核心功能：**状态推演预测、反事实推理、多方案并行模拟。
- ✧ **技术关键：**在隐空间中实现高效状态推演，无需显式渲染画面，通过抽象表征快速迭代实现认知推演，提升推演效率。

3.3.5 主动推理与行动规划闭环层（决策与行动）

作为最高决策中枢，它基于贝叶斯主动推理（Active Inference）等机制，将感知、意图与行动融为一体。它不仅自下而上地接收环境反馈，更自上而下地通过预测误差来指导底层感知并生成最优化行动路径，形成知行合一的闭环。对应人脑前额叶皮层、基底神经节与运动皮层，负责制定行动策略、输出控制指令，通过行动获取反馈更新世界模型，形成完整闭环。

- ✧ **核心功能：**目标导向规划、行动策略选择、闭环更新迭代。
- ✧ **技术关键：**基于主动推理框架，将感知、决策、行动与学习统一到同一生成模型中，以先验偏好或学习到的偏好表达目标，并通过期望自由能同时权衡目标实现与信息获取。

第四章 类脑智能全产业链协同创新发展建议与行动倡议

类脑认知世界模型是类脑智能的核心方向，其产业化落地覆盖算法理论、模型工程、芯片硬件、系统集成与基础支撑的全链路系统工程，而非单一技术的单点突破。只有构建层级协同、深度融合的完整技术体系，才能充分释放类脑认知的能效优势，支撑具身智能等场景的商业化落地。

当下，类脑智能产业面临的理论转化不足、技术卡点突出、数据供给薄弱、标准体系缺失、产业生态割裂等核心问题，要坚持问题导向与实效导向，围绕“补短板、建体系、促协同、推落地”总体思路，推动形成“政府引导、科研驱动、企业主体、产业协同”的发展格局。因此，本白皮书基于对类脑智能产业从算法到芯片的全产业链路分析，向业界提出多项产学研协同倡议。

4.1 类脑认知世界模型从算法到芯片的产业链路

机制层是类脑认知世界模型的理论内核，通过感知编解码、认知动态预测、终身学习记忆、低功耗稀疏计算四大核心类脑机制，实现从“数据统计拟合”向“主动认知建构”的范式转变，从底层突破大模型高数据依赖、弱泛化能力、无终身学习、高运行功耗的结构性瓶颈。

模型层是类脑认知世界模型的工程化载体，采用类脑分层架构，通过感知编码、状态推演、策略规划、学习更新四大模块，形成“预测-误差-行动-更新”的闭环运行机制，成为具身智能体的通用认知计算引擎。

计算层是能效落地的重要基础。传统 CPU/GPU/NPU 可通过稀疏算子、量化、缓存与异构计算支持类脑算法；存算一体、事件驱动加速器和神经形态/SNN 芯片则提供更专用的硬件路径。不同方案的能效可在相同任务、精度和时延约束下比较端到端能耗。

系统层实现类脑认知世界模型与终端硬件本体的深度融合，构建“感知-认知-控制-执行”分层架构与“端侧实时决策+云端聚合进化”的端云协同体系，打通从认知决策到物理执行的完整链路。

基础设施层是产业规模化发展的基础保障，重点建设多模态物理交互数据集、高保真仿真训练平台与标准化评测体系，补齐数据、工具、标准三大产业短板。

整体来看，类脑认知世界模型遵循“机制为核、模型为体、芯片为基、系统为用、支撑为翼”的层级逻辑，各层协同创新共同构成了从理论到落地的完整技术路径。

4.2 专项布局：推动类脑智能重大科技专项落地

遵循技术发展规律，分四个层级精准施策，避免资源分散与重复建设。

- ✧ **稳定支持基础研究：**聚焦主动推理、预测编码、脑启发计算、分层记忆等底层方向，依托产学研主体共建联合实验室，实行长期稳定资助，放宽短期产出考核，鼓励原创性探索。
- ✧ **集中突破核心技术：**依托“揭榜挂帅”等项目申报机制，重点突破类脑计算机制、算法-芯片协同设计、sim2real 虚实迁移等共性技术；布局存算一体、脉冲神经网络芯片，形成“突破一批、储备一批、布局一批”的迭代格局。
- ✧ **推进场景示范落地：**聚焦工业机器人、智能制造、商业服务、康养养老等刚需赛道，采用“用户出题、企业答题、科研支撑”模式，开展落地验证与规模示范，以真实场景牵引技术迭代。
- ✧ **建立风险包容机制：**对类脑智能的前沿探索类项目设置合理失败容忍度，设立技术探索风险补偿资金，鼓励主体勇闯技术无人区。

4.3 数据基建：建设国家级物理智能开放数据体系

破解物理交互数据匮乏瓶颈，构建“基础公益+行业增值”的数据供给体系。

- ✧ **建设分级分类数据集：**国家牵头建设覆盖基础物理交互的多模态共性数据集，面向全行业公益开放；龙头企业牵头建设细分行业场景库，按市场化原则共享。
- ✧ **畅通数据流通机制：**统一数据采集、标注与格式标准。建立分级开放与合规交易机制，推动数据要素有序流通，平衡数据安全与产业效率。
- ✧ **多渠道扩充数据供给：**依托生成式数据仿真平台、地方数据采集中心、教育机构实训基地等多渠道批量生成交互数据，多维度满足多样性的数据场景覆盖。

4.4 标准先行：加快行业评测标准与规范体系建设

加快构建覆盖术语、评测、接口、安全的标准体系，规范产业发展秩序。

- ✧ **统一基础术语标准**：明确类脑认知世界模型的定义、分类与能力边界，发布技术路线图，纠正概念模糊泛化，凝聚行业共识。
- ✧ **建立评测认证体系**：围绕感知理解、动作条件预测、因果干预、开放环境泛化、持续学习、规划控制、运行能效与安全可靠等维度，建立统一基准与第三方评测制度；核心指标包括多步预测误差与校准、任务成功率、少样本适配曲线、遗忘率、

跨场景/跨本体泛化、端到端时延和单位成功任务能耗，并通过消融实验量化关键机制的独立贡献。

- ✧ **规范接口互联标准**：统一模型与传感器、执行器、端云协同的接口协议，降低系统集成成本，促进生态兼容。
- ✧ **完善安全伦理规范**：制定功能安全、信息安全标准与伦理指引，明确应用边界，保障技术健康可控发展。

4.5 开放合作：构建产学研用一体化产业共同体

推动多主体协同联动，构建产学研用贯通、上下游联动的创新生态。

- ✧ **深化产学研协同**：共建联合实验室与技术创新中心，推行“企业出题、高校解题、市场验证”模式，联合培养交叉学科人才，促进成果转化。
- ✧ **强化产业链协同**：推动算法企业、芯片企业、机器人厂商、行业用户联合研发与适配验证，打通从技术到应用的全链条。
- ✧ **促进大中小企业融通**：龙头企业开放技术底座与工具链，专精特新企业深耕细分领域，形成梯度发展、融通共生的生态。
- ✧ **拓展国际合作**：跟踪全球前沿技术进展，吸引高端人才与团队，参与国际标准制定，深度融入全球创新网络。

第五章 总结与展望

5.1 核心总结

第一，世界模型是推动具身智能进阶、迈向通用智能的核心技术方向之一；类脑智能路线为突破数据效率、场景泛化、持续学习、能效四大瓶颈，提供了兼具理论根基与落地潜力的高价值路径。

当前以大语言模型为核心的深度学习范式已显现结构性瓶颈，难以支撑具身智能时代开放环境、长期自主、实时交互的核心需求。世界模型通过构建环境的内部表征与推演能力，成为迈向通用人工智能与具身智能的必经之路。全球技术演进趋势表明，单纯依靠数据规模与算力堆砌的工程化路径已接近增长天花板，向脑科学寻求底层机理、走类脑智能发展道路，是突破当前范式瓶颈的根本方向，代表了下一代 AI 的长期演进趋势。

第二，类脑认知世界模型是面向行动的认知推演系统，具备差异化的核心能力优势。

类脑认知世界模型以主动推理、预测编码为核心原理，采用“感知抽象-因果表征-时空记忆-反事实推演-行动闭环”五层技术架构，通过四大类脑认知计算机制，实现了从“数据拟合”到“认知建构”的范式转变。相比大模型驱动路线，其在小样本学习、跨场景泛化、终身持续学习、端侧低功耗等方面具备显著优势，能够从底层破解当前具身智能“高数据依赖、低泛化能力、灾难性遗忘、高运行功耗”四大困境，更适配开放物理世界的具身智能需求。

第三，全球产业呈现多流派分层演进、加速融合的态势，我国具备良好发展基础与战略机遇。

当前全球世界模型形成五大技术流派，分处不同技术层级，整体呈现“底层感知工程化提速、上层认知类脑化收敛”的演进趋势，未来将走向“认知为核、感知为用”的深度融合。我国具备超大规模应用市场、完备制造业体系、政策持续加码、区域产业集群快速形成等优势，在类脑智能赛道与全球领先水平差距较小，具备技术追赶与产业突围的良好基础，正处于抢占下一代 AI 技术制高点的关键窗口期。

第四，产业发展需坚持理论驱动与工程落地并重，走产学研协同创新道路。

类脑智能产业仍处于发展早期，面临理论转化难、核心人才缺、数据供给不足、标准生态缺失等多重挑战。其发展不能延续大模型时代“重工程、轻理论”的路径依赖，必须坚持第一性原理，以脑科学理论为引领，以工程化落地为抓手，构建产学研用深度协同的创新体系，走难而正确的长期发展道路。

5.2 发展展望

5.2.1 技术演进展望

短期来看（未来 3-5 年），模型侧，类脑认知世界模型核心算法将逐步成熟，四大核心计算机制实现关键突破，小样本学习、因果推理、终身学习等核心能力指标达到实用水平；**计算侧**，模型-芯片协同设计取得重要进展，存算一体、脉冲神经网络等类脑芯片实现规模化应用，端侧部署的性能与功耗达到商业化要求；**应用侧，**空间建模、视频生成、仿真交互等技术与类脑认知框架深度融合，形成体系化的完整技术方案。

中长期来看（未来 5-10 年），类脑认知世界模型将实现从专用场景模型向通用认知模型的演进，具备跨领域通用认知、复杂任务自主规划、开放环境终身进化的能力。脑科学、人工智能、类脑芯片、机器人学等多学科深度融合，催生新一代智能计算范式，推动人工智能从“感知智能时代”真正迈入“认知智能时代”。

5.2.2 产业发展展望

短期来看（未来 3-5 年），类脑认知世界模型将率先在工业制造、特种作业、商业服务等刚需场景实现规模化落地，形成一批标杆应用与龙头企业。同时，行业标准体系逐步建立，数据、算力、评测等公共基础设施日趋完善，产业生态初步形成。

中长期来看（未来 5-10 年），技术成熟度持续提升，应用场景向家庭服务机器人、无人机、自动驾驶、消费电子等领域广泛渗透，催生万亿级产业规模。最终形成覆盖算法、芯片、系统、应用的完整产业链，构建自主可控的产业生态，我国将成为全球类脑认知世界模型的重要创新中心与产业高地。

5.3 类脑智能根技术创新的战略意义

站在中美竞争、全球技术发展最前沿的角度，中国在计算机科学、脑科学、类脑产业、量子计算、核聚变等诸多领域实现了从追赶到并跑到局部领跑的阶段。类脑认知世界模型作为下一代人工智能的核心技术，将为我国在全球 AI 根技术创新、构建全球 AI 第二极技术生态版图的全球博弈中，抢占长远战略制高点、筑牢底层发展根基。

《类脑认知世界模型技术白皮书》的发布将为上海打造全球类脑智能创新策源地、抢占未来产业制高点提供关键支撑。更将成为我国人工智能领域高水平科技自立自强的重要抓手。突破类脑智能核心技术，构建自主可控的技术体系与产业生态，不仅将推动我国人工智能技术实现从跟跑到并跑、领跑的跨越，更将赋能制造业、服务业等实体经济智能化

升级，培育新质生产力，为建设科技强国、数字中国提供核心技术支撑。

参考文献：

1. A Path Towards Autonomous Machine Intelligence
2. Moravec, H. Mind Children: The Future of Robot and Human Intelligence. Harvard University Press, 1988.
3. Friston, K. J. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11, 127–138, 2010
4. Parr, T., Pezzulo, G., & Friston, K. J. Active Inference: The Free Energy Principle in Mind, Brain, and Behavior. MIT Press, 2022.
5. Rao, R. P. N., & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience, 2, 79–87, 1999
6. The Nature of Explanation（《解释的本质》），Kenneth J. W. Craik | 1943 年
7. Cognitive maps in rats and men,Edward C. Tolman | 1948 年
8. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely moving rat,O'Keefe, J.、Dostrovsky, J. | 1971 年
9. A Path Towards Autonomous Machine Intelligence,Yann LeCun | 2022 年
10. Assran, M. et al. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. CVPR, 15619-15629, 2023.
11. Bardes, A. et al. Revisiting Feature Prediction for Learning Visual Representations from Video. arXiv:2404.08471, 2024.
12. Lanillos, P. et al. Active Inference in Robotics and Artificial Agents: Survey and Challenges. arXiv:2112.01871, 2021.
13. Mumford, D. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biological Cybernetics, 66, 241-251, 1992.
14. Bruce, J. et al. Genie: Generative Interactive Environments. arXiv:2402.15391, 2024.
15. V-JEPA: Video Joint Embedding Predictive Architecture,Meta AI | 2024 年
16. Active Inference for Intelligent Agents: A Technical Overview,Verses AI | 2022 年
17. Free Energy Principle for Embodied Robotics,Verses AI | 2024 年
18. Genie: Generative Interactive Environments,Google DeepMind | 2025 年
19. Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. Active Inference: Demystified and Compared. Neural Computation, 33, 674-712, 2021
20. 《“中国脑计划”（脑科学与类脑研究）总体发展规划》 ，科技部等国家部委 | 2021 年

21. 《中华人民共和国国民经济和社会发展第十五个五年规划纲要》，全国人民代表大会 | 2026 年
22. 《关于可解释、可通用的下一代人工智能方法重大研究计划》，国家自然科学基金委 | 2025 年
23. 类脑智能是下一代 AI 的重要突破方向 https://www.ncsti.gov.cn/kjdt/kjrd/202606/t20260604_248803.html
24. 2026 类脑智能产业创新发展大会在沪成功举办，点亮未来产业新图景 https://baijiahao.baidu.com/s?id=1866062026324576322&wfr=spider&for=pc
25. 依旧少年丨蒲慕明院士：为中国脑计划“定心”，在“大脑竞技场”跑出中国速度_新浪财经_新浪网 https://finance.sina.com.cn/stock/t/2026-09-02/doc-iniqmmyy7714782.shtml
26. “AI 教母”李飞飞深夜发文：所有人都在造“世界模型”，但根本不是同一件事-智源社区 https://hub.baai.ac.cn/view/55240

