

附件 5

“人工智能的数理基础”重点专项 2026 年度项目申报指南 (征求意见稿)

为落实“十五五”期间国家科技创新有关部署安排，国家重点研发计划启动实施“人工智能的数理基础”重点专项。根据本重点专项“十五五”实施方案的安排，现发布 2026 年度项目申报指南。

“人工智能的数理基础”重点专项总体目标是：聚焦深度学习与大模型的基础理论、认知功能的数学建模、跨方法融合的关键问题以及可验证、可交互、可落地的系统建设，推动面向复杂场景的系统能力验证，在人工智能原型系统等方面取得突破，力争形成具有我国自主性的理论体系与方法论。

2026 年度指南围绕深度学习和大模型基础理论、物理智能、数学智能与格物智能逻辑环境、认知功能基本原理与原型系统构建等四个重点任务部署项目，拟支持 15 个常规项目，拟安排国拨经费概算约 2.647 亿元；拟支持 26 个青年科学家项目，拟安排国拨经费概算约 0.78 亿元，每个项目 300 万元。

项目统一按指南二级标题（如 1.1）的研究方向申报。申报项目的研究内容必须涵盖二级标题下指南所列的全部研究

内容和考核指标，项目一般实施周期为 3 年。除指南中特殊说明外，每个指南任务原则上拟支持项目数为 1 项。常规项目下设课题数原则上不超过 4 个，项目参与单位总数不超过 6 家。常规项目设 1 名负责人，其下每个课题设 1 名负责人。本专项鼓励青年科研人员作为项目负责人或项目主要骨干参与常规项目申报。

青年科学家项目不再下设课题，项目参与单位总数不超过 3 家。项目设 1 名项目负责人，男性应为 38 周岁以下（1988 年 1 月 1 日以后出生），女性应为 40 周岁以下（1986 年 1 月 1 日以后出生），原则上团队其他参与人员年龄要求同上。除指南中特殊说明外，每个指南任务原则上拟支持项目数为 1 项。

1. 深度学习和大模型基础理论

1.1 大模型低秩表达驱动的架构设计、高效训练与国产芯片适配（青年科学家项目）

研究内容：围绕大模型权重、梯度、注意力矩阵/张量中普遍存在的低秩特性，研究低秩驱动下大模型架构与表达能力的统一理论，刻画秩约束对注意力、前馈、嵌入等不同架构组件的函数表达能力、泛化性能与收敛行为的影响规律，形成“秩-网络架构-表达能力”的整体关联图谱与理论；发展泛化性能与训练收敛双约束下的秩下界理论，定量刻画秩规模与表达能力的映射关系，并利用低秩特性指导网络结构设

计，提出低秩感知的大模型架构设计原则与加速训练方法；面向国产 GPU 与异构计算芯片，研究低秩矩阵/张量运算与大模型预训练、后训练及微调的软硬件协同优化方法，研制低秩加速训练算法与通用算子库，并在典型大模型任务中验证实际训练加速效果。

考核指标：大模型低秩表达理论覆盖全连接层、注意力计算、前馈网络等不少于 3 类核心运算，给出低秩近似误差上界与低秩网络架构的加速比量化公式。以 Qwen3.5-4B 等同规模开源稠密模型为基准，推导精度损失不超过 1% 的秩下界估计方法，据此指导低秩网络结构设计，并在不少于 5 种公开基准数据集上验证训练精度与加速比，理论预测与实验相对偏差不超过 10%。研制定配不少于 2 种国产 GPU 或异构芯片的低秩算子库，覆盖低秩分解、低秩矩阵乘等核心算子；以芯片厂商官方同精度全秩稠密算子库为基线，矩阵/张量计算核心环节加速比不低于 1.5 倍；以 Qwen3.5-4B 等同规模开源稠密模型的全秩训练为基线，端到端训练提速不低于 15%，核心精度指标损失控制在 1% 以内。在国产 AI 加速芯片上的大模型训练典型应用场景技术就绪度（TRL）达到 4 级，构建完整低秩算子库与训练系统并在国产芯片集群环境完成验证。研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：大模型表达能力，低秩近似，国产芯片适配。

经费说明：中央财政资金 300 万元。

1.2 面向推理任务的大模型表达理论与架构设计（青年科学家项目）

研究内容：针对大模型在有限长度推理任务上的表达能力，以推理长度作为约束，研究保证大模型正确推理所需的思维链长度、模型参数量以及参数精度的紧致上、下界理论，推导融合此三类要素的大模型推理资源需求；针对大模型解决长程推理任务时存在的上下文遗忘、思维链断裂难题，研究基于多模型协作推理的表达理论，构建大模型层次化分解长推理问题的理论机理，发展面向长距离依赖的分块求解、结果聚合算法；依托大模型表达理论和多模型协作表达理论构建推理智能体架构，研发高算力资源利用率与具备长程推理能力的推理智能体，面向 2 类典型推理语言形成稳定可靠的推理流程，在解决数学猜想与证明复杂定理中取得突破。

考核指标：构建 1 套涵盖思维链长度、模型参数量与参数精度的大模型推理资源需求理论模型；提出 1 套多模型协作推理表达理论，构建长推理问题的层次化分解机理，形成 1 套面向长距离依赖的大模型协同算法，在长程推理任务场景下，相比于标准 MoE 范式的任务准确率提升 30%；构建 1 套具备长程推理能力的智能体架构，相比于标准 ReAct 架构智能体的最大推理长度提升 30%，长推理任务上的推理准确率提升 15%，可兼容自然语言推理与形式化语言推理，在 2 个

数学猜想或复杂定理证明中取得显著突破。在数学猜想证明等典型推理场景技术就绪度（TRL）达到 4 级，构建完整推理智能体并在实验室环境完成验证。研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：推理任务表达能力，协作表达理论，推理智能体。

经费说明：中央财政资金 300 万元。

1.3 预训练尺度律的形成机理与调控（青年科学家项目）

研究内容：针对大模型预训练尺度律形成机理尚未明晰的问题，利用随机矩阵理论建立简化模型与大模型预训练尺度律之间的定量对应理论，刻画预训练损失演化规律；分析数据谱结构、模型容量与训练过程对尺度律的共同作用，揭示尺度律指数、有限规模效应及失效边界的形成机理；探索尺度律形成机制与训练临界稳定性的内在联系，形成面向大模型预训练稳定性的调控方法。

考核指标：建立基于随机矩阵理论的预训练尺度律与临界动力学理论，揭示数据分布、特征结构和曲率谱特征对尺度律形成、演化及失效的影响机制，定量解释不少于两种典型预训练现象；在参数规模不低于 20 亿参数的开源语言模型上完成实验验证，实现对训练稳定边界和异常动力学行为的预测，并使训练发散、损失异常振荡等不稳定现象发生率较代表性基准配置下 Adam 算法降低不低于 10%。在大模型预

训练技术验证系统技术就绪度（TRL）达到 4 级，构建完整系统并在实验室环境完成验证。研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：尺度律，随机矩阵理论，临界动力学，边缘稳定性。

经费说明：中央财政资金 300 万元。

1.4 大语言模型后训练的基础理论与算法研究（青年科学家项目）

研究内容：针对大语言模型后训练中反馈成本高、复杂偏好难以刻画等问题，研究主动反馈采样、偏好建模与偏好优化理论。探索思维链增强长程推理能力的作用机制，设计面向长上下文与多轮决策的后训练理论与算法，分析长轨迹训练和反馈延迟对训练稳定性与任务完成率的影响，提高多轮决策优化效率。研究面向多智能体的协同优化、交互训练信号与责任归因机制，发展高效、可归责的多智能体后训练方法。探究后训练持续学习中的参数更新与知识记忆的基本规律，分析新知识适应与原有能力保持之间的关系，研究缓解灾难性遗忘的持续学习方法。

考核指标：建立主动反馈采样与偏好优化的复杂度分析理论框架，提出降低反馈需求的后训练算法，以经典后训练方法（例如 PPO、DPO）为基线，在相同反馈样本量下，偏好胜率或任务成功率相对提升不低于 10%。形成面向长上下

文与多轮决策的后训练算法，长程任务完成率或多轮决策性能相对提升不低于 10%。分析多智能体协同优化的计算复杂度和责任归因可靠性，发展可归责的协同后训练方法，以独立智能体优化或简单协作策略为对比，协作任务成功率和责任归因准确率相对提升不低于 10%。刻画持续学习中的稳定性-可塑性关系，提出缓解灾难性遗忘的持续学习算法，相比推理增强持续学习方法，任务序列平均准确率相对提升不低于 10%。在大语言模型后训练典型应用场景的技术就绪度（TRL）达到 4 级，构建完整后训练方法系统并在实验室环境中完成验证，研究形成的高水平论文不少于 2 篇发表在我国科技期刊上。

关键词：模型后训练，偏好学习，长程决策，持续学习，多智能体协同。

经费说明：中央财政资金 300 万元。

1.5 面向国产算力的混合精度误差传播分析与混合精度算子研究（常规项目类）

研究内容：针对混合精度条件下神经网络前向与反向传播中的误差累积问题，研究有限精度、量化表示及不同计算实现方式引发的误差传播规律，构建计算顺序、分块方式和数据格式影响的误差传播数学模型。基于网络误差敏感度分析理论，构建层级误差贡献度量模型和误差传播图，识别影响模型输出质量和梯度估计精度的关键层、关键算子与关键

路径，推导重点算子的精度需求和误差容忍边界。面向国产 GPU、NPU 等 AI 芯片架构，研究误差敏感算子的混合精度计算重构方法，形成“误差传播模型-精度需求推导-混合精度算子设计-国产硬件实现”的一体化框架。

考核指标：构建混合精度前向与反向传播误差传播数学模型，覆盖矩阵乘法、注意力机制、残差连接、MOE 等不少于 6 类关键模块和 transformer 整体机构。构建整网误差敏感度分析方法，识别对模型输出质量和梯度估计影响显著的关键层、关键算子和关键计算路径，理论分析和实验测试结果的前 30 个误差敏感模块重叠率不低于 70%。推导重点混合精度算子的精度需求和误差容忍边界，研发不少于 10 个适配国产 AI 芯片的混合精度高效算子。在参数规模超过 200B 的主流大模型（如 Qwen、DeepSeek）训练与推理测试中，在使用相同数据集及相同训练、推理算法的条件下，确保模型主要任务指标相较于默认精度方案的下降幅度不超过 2%、梯度方向一致性不低于 0.9 的前提下，实现显存占用降低不低于 30%，关键算子效率提升不低于 1.4 倍，端到端效率提升不低于 15%。技术成熟度（TRL）达到 4 级，在实验室环境中完成混合精度训练验证。研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：混合精度计算，误差传播，混合精度算子，国产算力。

经费说明：中央财政资金 1100 万元。

1.6 生成模型统一动力学与随机控制方法（青年科学家项目）

研究内容：围绕自回归、扩散、跳跃生成、概率流和流匹配等生成方法在动力学描述、训练机制和采样策略方面缺乏统一理论框架的问题，研究生成模型的统一动力学与随机控制方法。研究离散与连续、随机与确定性生成机制的统一表示及相互转换，建立生成模型统一动力学理论与算法框架；研究生成路径、噪声调度和时间推进等动力学因素对生成质量与效率的影响，发展高效生成与动力学调控方法；针对高维多峰分布中的亚稳态、跨模态转移困难及模式覆盖不足等问题，研究其动力学机制并发展相应调控方法；进一步从随机控制角度研究生成模型训练与采样，建立生成质量、稳定性、计算效率与控制代价之间的理论关系，形成生成动力学与随机控制相结合的方法体系。

考核指标：建立生成模型统一动力学理论框架，覆盖不少于 2 类典型状态空间，统一描述不少于 3 类生成机制，形成不同生成机制的统一表示或转换方法；形成不少于 2 类动力学调控或高效生成方法，代表性任务中生成计算效率提升 30% 以上；形成不少于 2 类面向多峰分布的生成调控方法，显著提升跨模态转移或稀有模态覆盖能力；形成不少于 2 类随机控制驱动的生成训练或采样方法，在可达性、稳定性或

收敛性等方面取得理论结果。完成不少于 3 类典型生成场景验证，覆盖不少于 2 类异构数据类型；形成算法原型及开源代码或可复现实验平台，典型应用场景技术就绪度（TRL）达到 4 级。研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：生成模型，统一动力学，生成加速，多峰生成，随机控制。

经费说明：中央财政资金 300 万元。

1.7 神经网络增强的高维概率分布密度估计与蒙特卡洛采样方法（常规项目类）

研究内容：面向 AI 前沿领域高维数据分布拟合与蒙特卡洛采样的精度、效率及稳定性瓶颈，研究高维分布建模与高效采样方法。研究结合神经网络的高维概率分布密度估计方法及其蒙特卡洛采样方法和理论，构建具有优良估计精度和严格理论保证的高维密度估计方法体系，突破高维数据分布拟合核心瓶颈；针对神经网络驱动的概率生成框架相结合，设计更有效的序贯蒙特卡洛方法，提升实际生成质量与计算效率；研究将序贯蒙特卡洛方法应用于大语言模型推理的方法和理论，提高大语言模型解空间效率，提高复杂推理任务的可靠性；面向具身智能领域中的深度强化学习问题，以及生成式人工智能中在复杂条件下的高质量条件生成问题，研究新理论、新方法的落地应用路径，推动前沿人工智能方法

在重大应用问题中取得新突破。

考核指标: 1 个将神经网络驱动的概率分布密度估计新兴方法与非参数统计经典理论方法有效融合的理论框架, 以及至少 2 种神经网络增强高维概率分布密度估计新方法, 针对具有复杂结构的高维概率分布的概率密度估计精度提升至少 20%; 1 个将神经网络驱动的生成式采样方法与蒙特卡洛采样经典方法有效融合的理论框架, 以及至少 2 种神经网络增强高维概率分布蒙特卡洛采样新方法, 针对具有复杂结构的高维概率分布的采样效率提升至少 20%; 针对具身智能领域中的深度强化学习问题和生成式人工智能中复杂条件下的高质量条件生成问题, 各自发展至少 1 种新方法, 综合运用上述新型计算技术和计算框架, 将深度强化学习和条件生成的综合效率提高至少 20%。

关键词: 高维密度估计, 高维分布采样, 预训练神经网络。

经费说明: 中央财政资金 1160 万元。

1.8 深度学习泛化与标度律的统计物理理论 (常规项目类)

研究内容: 面向大模型在预训练、后训练与推理全过程中呈现的泛化、标度律与能力涌现等典型现象与规律性问题, 研究深度学习泛化与标度律的统计物理理论。揭示过参数化模型良好泛化与隐式正则化的物理机理, 研究预训练表征学

习、能量景观与下游任务行为之间的内在关联，构建深度学习泛化的统计物理理论体系；研究数据内在结构、模型架构、训练算法与算力规模共同决定幂律标度的微观机制，阐明守恒律、对称性与多尺度结构等物理约束对文本、图像及分子、材料、蛋白等数据模态标度律的影响，发展标度律的第一性原理定量预测方法并界定其适用范围；发展连续极限、热力学极限与重正化群等统计物理方法，构建由少数宏观量支配的有效理论，提出刻画表征结构与相变行为的关键物理量，阐明能力涌现的临界条件；建立“泛化机制解析-标度律定量预测-涌现临界判定-资源配置指导”的理论与方法体系；在语言、视觉与多模态大模型及科学基座模型等典型系统上开展验证，检验理论与方法的普适性。

考核指标：建立 1 套深度学习泛化与标度律的统计物理理论框架，系统解释泛化、标度律与能力涌现；提出不少于 2 个可检验的定量预测模型，对过参数化模型良好泛化、标度律、能力涌现、表征相变等不少于 4 种现象给出定量解释，在典型模型或公开基准上完成验证。验证基于真实数据与真实规模算力，覆盖 10 亿量级以上参数模型，模型规模、数据量与算力跨越至少 3 个数量级；在文本、图像与分子、材料、蛋白等不少于 3 类数据模态上，实现标度律第一性原理定量预测，幂律指数预测相对误差不超过 15%，更大规模外推精度优于 Chinchilla 等经验拟合基线。形成可复现的预测方法与

工具，提出不少于 2 条用于规模扩展、资源配置与能力预测的原理性指导，支撑科学基座模型构建；技术就绪度（TRL）达 4 级，代表性论文发表在我国科技期刊上的比例不低于 20%。

关键词：统计物理，标度律，泛化理论，重正化群，科学数据。

经费说明：中央财政资金 1160 万元。

1.9 神经网络训练的物理模型与优化（青年科学家项目）

研究内容：从物理基本原理出发，分析理解神经网络的训练过程。运用统计物理、非平衡动力学与随机过程等物理概念和方法，系统分析训练过程的动力学特征与典型行为，提炼刻画训练过程的关键物理量，揭示训练过程中普适性规律。建立神经网络训练过程的物理模型，定量描述与分析训练行为。将神经网络训练刻画为高维能量景观上的非平衡动力学过程，构建可定量计算的物理模型，发展学习过程的响应理论等理论分析方法，并用不同任务、不同规模的真实神经网络训练数据，定量检验模型预测的准确性与适用范围。依据物理模型揭示的能量景观等规律，提出训练过程应遵循的定量设计规则，并据此设计新的训练优化算法，形成可工程实现的优化方法，并验证优化方法且开源实现，在语言模型、视觉与多模态模型、深度生成模型等主流神经网络任务上，定量验证所提优化方法相对 Adam 优化器在训练效率与模型性能上的提升幅度；在电子结构计算、原子间相互作用

建模等科学神经网络典型任务上检验方法的适用性与泛化能力。

考核指标:建立神经网络训练的非平衡动力学理论框架,提出可定量描述学习过程的物理模型,其对学习曲线与训练相变的预测在真实神经网络训练实验中得到定量验证;提出基于物理模型的神经网络训练优化方法,在语言模型、视觉或深度生成模型等主流神经网络及电子结构神经网络、机器学习原子势等科学神经网络上,相对于 Adam 优化器实现不少于 30%的训练效率提升(达到同等模型性能所需训练算力降低 30%以上),或同等算力预算下模型性能的显著提升(科学神经网络任务中预测精度提升 30%以上),验证任务不少于 3 类。所建立的神经网络训练物理模型与优化方法技术就绪度(TRL)达到 4 级,构建开源代码原型并在实验室环境完成验证。研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词: 神经网络训练, 非平衡动力学, 统计物理, 优化算法。

经费说明: 中央财政资金 300 万元。

1.10 基于不确定性量化的智能算法内生安全理论与方法(青年科学家项目)

研究内容:以大模型为代表的智能算法在数据-模型-决策全链路中均存在不确定性,是其内生安全风险的重要来源。

研究模型在数据不完备情景下采样偏差的误差界，构建基于数据不确定性量化的对抗防御机制；研究不同模型结构和优化训练算法下模型分布外泛化界，构建高效鲁棒优化算法和最优对抗鲁棒建模策略；研究模型决策置信度的度量方法，突破置信度校准和多源融合可信决策难题。面向互联网内容安全治理或其他典型场景，应用所发展的方法构建安全可信大模型或智能体系统，实现在动态开放环境下应对对抗攻击、分布偏移、多源决策融合的能力提升。

考核指标：构建数据不确定性下的对抗误差界、模型不确定性下的分布外泛化界，智能学习算法内生安全增强方法；构建模型决策置信度度量以及多源融合的可信决策方法；构建互联网内容安全治理或其他典型应用场景的大模型或智能体：在至少 3 个典型风险内容识别数据集上，越狱攻击的防御成功率与对抗训练方法相比提升 15%以上，置信度校准误差相对降低至少 30%，所提模型可信融合决策推理成功率提升 5%以上。在互联网内容安全治理或其他典型应用场景技术就绪度（TRL）达到 4 级，构建完整系统并在实验室环境完成验证。研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：不确定性量化，对抗鲁棒性，分布外泛化，安全可信决策。

经费说明：中央财政资金 300 万元。

2. 物理环境和物理智能

2.1 融合多方法的空间感知、理解、建模与验证环境（常规项目类）

研究内容：研究几何（如基于 CAD）、逻辑、物理（如基于微分方程）与数据方法深度融合的新方法，发展新的空间智能建模、理解与推理技术。研究融合几何约束、守恒关系、物理规律和逻辑规则的几何物理属性估计、状态演化预测与因果推理方法，并构建相应验证环境。面向未见场景、稀有事件、观测缺失和长时程推演，研究基本原理驱动的泛化与纠错机制。在实验室、工业产线、家居环境、物流环境等场景落地实现。在同等数据规模、模型规模和评测任务条件下，验证融合方法在建模误差、理解准确率和推理精度方面整体优于纯数据驱动基线，且建模和推理结果必须满足几何与物理属性约束。

考核指标：构建多方法融合机制，复杂场景下调度成功率不低于 95%，处理效率不低于同等算力和任务条件下的纯数据驱动方法。主要对象采用 CAD、参数化几何、拓扑、运动副或功能结构方法进行可计算表示；建立不少于 3 类可检验几何约束，建模结果必须符合几何要求。对纯数据生成的不合理交互对象进行几何、物理和逻辑校正，检出率不低于 95%，校正后符合几何、物理和逻辑要求。构建 1 套新方法验证环境，支持不少于 6 类输入模态、4 类方法接入和 3 类具

体任务；形成不少于 1000 个标准测试样例，支持与纯数据驱动方法开展同条件对比、消融分析和结果复现。验收时达到技术就绪度 TRL 5 级，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：几何物理约束，空间智能建模与推理，融合方法验证环境。

经费说明：中央财政资金 2400 万元，自筹配套资金不低于 600 万元。

2.2 数据、逻辑、几何、物理融合驱动的物理世界认知与交互方法（青年科学家项目类）

研究内容：研究数据观测与任务逻辑、几何关系和物理规律协同的认知方法，根据任务目标确定需要认知的主要交互对象及关键属性，利用任务先验、几何关系和物理规律减少对真实观测数据的依赖，在较少数据条件下实现交互对象状态、属性和关系的有效认知。研究从真实世界多模态观测数据中反演主要交互对象几何结构、功能关系和关键物理属性的方法，使建模结果符合任务相关的几何要求、接触关系和物理规律。研究融合任务逻辑、空间关系和物理规律的动作后果推演、交互可行性判断与行为生成方法，使认知结论、操作方案和交互行为通过任务逻辑、空间可达性和物理可行性检验。基于国产算力平台，在家居、物流、工业等至少 1 类典型真实场景中完成不少于 3 类复杂交互任务验证。

考核指标：面向至少 1 类典型场景从真实观测数据中反演主要交互对象的几何结构和关键物理属性；可测量关键物理属性相对于实验测量值或标定参考值的平均相对误差小于 5%。在达到同等任务性能、且结果通过几何要求和物理规律检验的条件下，所需真实观测数据量较同规模纯数据驱动方法降低不低于 30%。认知结论、操作方案和交互行为应通过任务逻辑、空间可达性和物理可行性检验，且优于纯数据方法 50% 以上。在至少 1 类真实场景中完成不少于 3 类复杂交互任务，主要工况下任务成功率大于 95%；在任务书明确的国产算力平台、输入规模和并发条件下，端到端推理时延不高于 100 ms。验收时达到技术就绪度 TRL 5 级，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：任务逻辑牵引，物理属性反演，小数据认知，动作后果推演，交互可行性检验。

经费说明：中央财政资金 300 万元。

2.3 复杂生物环境的智能理解、动力学推理与交互决策方法（常规项目类）

研究内容：面向复杂连续生物组织在有限观测和实时交互条件下的状态理解、动力学推理与干预决策需求，研究非均质、多尺度及多介质耦合条件下的物理状态实时表征与动力学建模方法。研究融合多模态观测、高保真模拟数据与物理知识的生物组织结构、属性及动力学状态统一表征方法；

研究复杂介质动力学智能模型构建方法，针对生物组织组成的固-液多介质耦合体系，研究波动传播与力学响应的物理一致性建模及有限声学观测下的状态与参数反演，构建高保真数值模型与物理约束智能模型之间的映射关系，研究模型在线更新与因果干预推理方法，融合实时观测数据、力学约束残差、不确定性评估和误差反馈，实现模型对个体差异和外部干预的自适应修正；面向骨组织诊疗等物理/数字场景开展验证。

考核指标：构建复杂物理约束智能建模与推理方法 1 套，支持不少于 2 类典型含骨组织动力学过程的状态估计、演化预测与干预效应推理。能够预测复杂生物组织动力学状态与关键参数的变化规律，预测结果与实测变化趋势一致；关键力学响应预测误差相对于高保真数值结果不超过 2%，物理约束残差不超过 3%；典型动力学任务计算效率较传统高保真方法提升 5 倍以上，计算与实验平台 100%国产化。完成不少于 3 类典型复杂物理/数字场景验证，在 10 例以上含骨病例或等效实验样本进行验证，形成可行的标准化落地方案。验收时达到技术就绪度 TRL 6 级，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：智能计算力学，物理约束，脊柱生物力学，高泛化，超声反演。

经费说明：中央财政资金 3000 万元。

2.4 物理智能流体与流固耦合行为预测与验证（青年科学家项目）

研究内容：以高动态交互场景流体与流固耦合行为快速预测为牵引，建立流固耦合实时神经算子求解器。融合高速视频、可微渲染与可微动力学，反演结构三维运动与变形状态、流场及气动、材性与接触参数。研究流体-结构-接触耦合神经算子求解器，基于多工况训练实现跨速度、材性等快速预测。融合实时视频观测与神经算子预测，逐帧校正结构运动与旋转状态，在线更新气动参数及模型残差，实现低时延滚动预测。面向机器人打乒乓球、机器人踢足球、船舶动力学、机翼振动等典型场景，验证结构运动状态与交互结果预测的准确性、实时性。

考核指标：建成 1 套面向高速变化的流体与流固耦合行为智能预测原型系统，在国产芯片开展仿真，基于视频的结构三维运动与变形状态、流场及气动、材性与接触参数反演误差不超过 5%；基于国产芯片，关键物理量神经算子求解器预测误差不超过 5%；在国产芯片上，跨速度、材性等参数条件下神经算子求解器外推测试精度下降不超过 5%；滚动推演结构三维运动状态误差不超过 5%；形成不少于 1 套流体与流固耦合预测系统，完成机器人打乒乓球、机器人踢足球、船舶动力学、机翼振动等至少 2 类典型场景验证。验收时达到技术就绪度 TRL 6 级，研究形成的代表性论文发表在我国科

技期刊上的占 20%以上。

关键词：流体与流固耦合，行为预测，神经算子求解器。

经费说明：中央财政资金 300 万元。

2.5 面向物理智能交互的复杂介质数字环境构建与实时响应方法（常规项目类）

研究内容：面向智能体在复杂介质环境中的实时感知、预测与交互需求，研究复杂介质数字环境构建与实时响应方法。重点突破跨尺度映射、真实物理状态反演、高精度实时响应计算和在线校正等关键技术，提升数字环境的物理一致性、实时性和可用性。围绕复杂交互场景中的典型复杂介质对象（如软物质、颗粒物质、多孔介质、复合材料等），研究真实物理环境到数字环境的一致性映射与统一表征，形成几何结构、材料属性、微介观状态和动力学响应统一的状态表达，并建立融合物理建模和数据驱动的复杂介质等效动力学框架。研究面向智能体交互的数字环境在线校正与可信切换方法，建立模型参数、状态变量和响应误差的动态修正机制，持续提升数字环境与真实物理世界的一致性。构建典型验证环境，在工业、实验室、物流等场景中，面向典型复杂介质对象的操控、输运、加载与装配等交互任务，评估模型在不同介质条件下的适用边界与可信度，验证在线可信切换、响应一致性、实时交互能力和国产算力平台适配能力。

考核指标：建立覆盖不少于 3 类复杂介质的数字环境，

以及融合物理建模和数据驱动的复杂介质等效动力学框架，具备几何、材料和动力学状态统一表达能力，并在应用场景中进行验证，如工业、实验室、物流等，且需在国产算力平台上满足实时响应要求，端到端响应频率 $\geq 30\text{Hz}$ 。建立真实环境到数字环境的一致性映射与状态反演方法，实现数字环境在线校正，真实环境关键状态参数反演误差 $\leq 5\%$ 。建立不少于3类场景的模型适用边界评价方法，实现在线可信切换，误判率 $\leq 2\%$ 。构建典型物理交互验证环境，在工业、实验室、物流等典型场景下，支持典型复杂介质对象的操控、输运、加载与装配等不少于3类智能体交互验证。验收时达到技术就绪度TRL 6级，研究形成的代表性论文发表在我国科技期刊上的占20%以上。

关键词：复杂介质数字环境，物理智能交互，等效动力学，在线校正。

经费说明：中央财政资金3000万元。

2.6 视觉、几何、物理一致的全要素数字环境高效构建与交互（常规项目类）

研究内容：面向工业、实验室、物流、医疗等典型真实场景，研究视觉、几何、物理一致的全要素数字环境高效构建与交互方法。建立典型真实场景全要素物理复刻与高效构建的软硬件系统，贯通视觉几何重建、对象部件解析、物理属性反演、功能结构恢复和交互行为校正，构建与真实环境

一一对应、可训练、可交互、可复用的高精度物理数字环境。研究典型真实场景与数字环境的一一映射构建方法，实现真实世界与数字空间中几何对象和视觉外观的逐一对应；研究可动部件、关节约束和功能关系的识别重建，以及质量、惯量、摩擦、刚度、阻尼等物理属性的反演与绑定，使数字环境与真实世界具有一致的结构功能和物理响应。构建原型系统，统一组织几何、语义、材料、碰撞、关节、功能及传感等全要素信息，形成物理可执行、功能可操作、行为可复现的数字环境。研究与真实场景的快速适配，有效消除 sim2real gap。

考核指标：建设如工业、实验室、物流、医疗等不少于 3 类典型实景及其一一对应的物理数字环境，实现几何对象、视觉外观与功能结构的对应构建。统一组织几何、语义、材料、碰撞、关节、功能及传感等不少于 6 类要素，形成包含物理属性与功能关系的场景全要素物理化表达。构建多模态物理感知软硬件系统，捕捉视觉、力觉和触觉等模态，并解决时空同步与低延迟传输难题。场景资产自动处理率不低于 80%，环境构建效率较人工流程提升不低于 100%，形成可复用的高效场景构建流程。支持对象操作、关节运动、碰撞接触和传感观测等物理可执行交互，支撑不少于 3 类复杂机器人任务的训练与预部署。验收时达到技术就绪度 TRL 6 级，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：全要素物理数字环境，视觉-几何-物理一致，交互状态校正。

经费说明：中央财政资金 3000 万元。

2.7 微创手术智能辅助决策与物理数字环境建模关键技术研究（青年科学家项目类）

研究内容：以微创手术精准化、智能化临床需求为牵引，瞄准术中实时感知、决策与物理环境建模的共性难题，研究基于多模态大模型的手术场景智能理解方法，并构建与之协同的物理智能数字环境，为手术辅助决策提供可交互、可推演的仿真验证平台。具体包括：研究基于多模态大模型的术中组织感知方法，构建术中实时分割与深度估计框架，实现关键解剖结构的像素级语义识别与相对位置定位；构建数理嵌入空间下可交互的术中物理数字环境；研究手术进程感知与决策推理方法，构建手术质控评估与完成判定机制；以内镜手术为典型验证场景开展临床数据验证，评估系统在组织识别精度、决策推荐准确率及物理仿真保真度等方面的综合性能。

考核指标：镜下组织语义分割平均交并比($mIoU$) $\geq 75\%$ ，实时推理帧率 ≥ 25 FPS，识别精度接近高年资医师水平；术中实时三维重建空间分辨率 ≤ 1 mm，重建完整度 $\geq 85\%$ ，单帧重建时间 ≤ 1 s；物理数字环境支持至少 3 类物理属性的数理表征，支持器械-组织不少于 3 类的力学交互仿真；手术阶

段识别准确率 $\geq 90\%$ ，关键操作节点检出召回率 $\geq 85\%$ ，风险预警准确率 $\geq 90\%$ ；对抗攻击防御后模型在两种典型攻击下攻击成功率降低 $\geq 20\%$ ，临床问答准确率保持率 $\geq 95\%$ ；智能体系统在 ≥ 10 例真实手术案例上完成临床验证，辅助决策建议临床可接受率 $\geq 90\%$ 。验收时达到技术就绪度 TRL 6 级，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：微创手术，数字仿真重建，多模态大语言模型，智能辅助决策。

经费说明：中央财政资金 300 万元。

2.8 面向非保守非光滑交互机制的物理环境构建与状态演化（青年科学家项目类）

研究内容：面向精密装配、摩擦夹持与操控或人机接触等典型非保守非光滑交互任务，研究物理环境构建方法与状态演化机制，重点研究连续运动阶段、接触与约束切换事件以及能量耗散过程的统一表征与动态演化，构建能够真实复现上述典型任务动力学过程的物理数字环境。研究几何、材料、接触关系和边界条件的参数化表达，为上述物理数字环境提供真实任务对象及其相互作用关系的统一描述。研究接触与约束切换事件中的碰撞与分离、黏附与脱粘、黏着与滑移、卡滞与释放等关键过程的统一表示与转换机理及其能量耗散机制，揭示约束切换条件下的状态演化规律，实现几何、运动和力学状态协同更新。研究融合少量真实数据的环境参

数校正方法与高效实时响应计算方法，构建能够实时复现真实任务动力学过程的数字环境原型，使关键运动和接触响应与真实任务过程保持一致。

考核指标：聚焦工业、物流、家居等 1 类典型场景中的复杂交互任务，至少在 1 类场景中构建 1 套物理数字环境，包含几何、材料、接触关系和边界条件的参数化表达模型，覆盖任务所需主要工况。在交互任务中支持碰撞与分离、黏附与脱粘、黏着与滑移、卡滞与释放等不少于 3 类关键过程的状态转换，涵盖约束切换触发条件与能量耗散机制，状态更新满足几何、运动和力学一致性。环境参数经少量真实数据校正后，关键运动和接触响应相对于实验测量或标定参考的平均相对误差不超过 3%。在任务书明确的国产算力平台、模型自由度、活动接触数量和求解精度条件下，状态更新频率不低于 200 Hz，连续运行不少于 10^4 步不发生数值发散。验收时达到技术就绪度 TRL 5 级，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：非保守非光滑交互，物理数字环境构建，状态演化与表征，接触与约束切换，参数校正。

经费说明：中央财政资金 300 万元。

2.9 面向高频交互的物理智能环境验证与智能体自进化技术（常规项目类）

研究内容：研究可扩展的标准化接入与任务评测；面向

不同智能硬件、智能体及感知决策模型，研究统一接口规范、模块化适配工具和示范流程，建立任务描述、环境配置、过程记录、结果复现和指标自动计算机制；研究动态任务生成、参数可控配置以及扰动与失效注入方法；构建动态、可控、可重复的高频交互验证场景和智能体全过程闭环验证；贯通环境感知、状态估计与预测、任务规划、动作执行和物理反馈，研究多源状态同步、不同频率闭环协同及交互过程结构化记录方法，形成面向任务性能、物理一致性、实时性、安全性和重复性的多维验证方法。开展仿真环境与实物系统的同任务对照测试，量化关键状态响应和任务性能的虚实差异，为环境参数校正和策略迁移提供依据、失败归因与自进化验证；研究覆盖感知、建模、规划、控制和环境交互等环节的失效表示、检测、定位、回放与归因方法，构建验证反馈驱动的环境修正、数据补充和策略优化机制，形成“验证-归因-优化-再验证”的低人工干预智能体自进化闭环，并支持未见任务、扰动条件和跨场景迁移能力的持续评测。

考核指标：研制标准化接入与模块化评测工具，支持验证任务自动结果复现和指标自动评估，形成可扩展、可重复的标准化验证流程；覆盖如工业、实验室、商超、家庭等不少于 4 类的典型场景，建设 30 项高频交互任务，接入不少于 4 类智能硬件、智能体或感知决策模型；在国产算力平台上，物理状态更新频率 ≥ 120 Hz，关键控制反馈频率 ≥ 60 Hz；相

同初始条件下重复验证结果一致率 $\geq 98\%$ ；覆盖感知、建模、规划、控制和环境交互等不少于 5 类失效；预置及回放失效的检测召回率 $\geq 90\%$ ，失效归因准确率 $\geq 85\%$ ；完成不少于 5 轮低人工干预的自主迭代，失效或未见任务成功率/关键性能提升 $\geq 20\%$ ；在不少于 2 类实物系统上验证，虚实任务性能差异 $\leq 10\%$ 。验收时达到技术就绪度 TRL 7 级，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：高频交互，物理智能环境验证，标准化接入，闭环评测，失效归因，智能体自进化，虚实迁移。

经费说明：中央财政资金 1500 万元，自筹配套资金 1500 万元。

2.10 基于复杂数字环境的物理智能基础能力与原理性验证（青年科学家项目类）

研究内容：面向多类物理智能系统，依托典型场景物理数字环境，构建覆盖环境感知与状态估计、物理属性与可交互结构判断、空间关系与动态规律推演、任务规划与交互控制的原理性通用智能验证体系。研究多类物理智能系统统一接入、能力-任务匹配和验证任务标准化描述方法；研究复杂动态场景、交互对象、状态演化和动作后果的结构化表示方法；研究物理智能基础能力分项验证、结果自动评价和能力画像生成方法。研究由验证结果和困难样例驱动的内容生成、质量筛选和体系持续更新方法；形成可接入任务 2 标

准化交付与验证平台的验证工具、任务库、实例库和演进案例库。

考核指标：建立综合能力的验证体系，形成不少于 30 项可独立测试的能力项目，涵盖空间感知、运动追踪、物理属性估计、交互推理等内容；支持空间智能体等物理智能系统接入；在如家居、实验室、物流等典型场景下，建设不少于 1000 个验证实例，覆盖不少于 50 种环境与物体配置、5 类物理对象和 15 类挑战条件；支持感知与状态预测结果自动评价，支持物理属性或结构参数回填仿真，覆盖多类动态交互过程和操作任务，自动评价覆盖率不低于 90%，重复测试结果一致率不低于 95%；建立验证体系持续演进机制，具备验证内容扩展与质量控制能力；研发验证平台及软件 1 套，建设测试场景库、评测套件和演进案例库。验收时达到技术就绪度 TRL 6 级，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：物理智能，原理性能力考核，物理数字环境验证，动态交互，验证体系自进化。

经费说明：中央财政资金 300 万元。

3. 面向数学智能与格物智能的逻辑环境

3.1 面向智能数学环境的可计算知识表达与多模态语料库构建（青年科学家项目类）

研究内容：针对数学知识来源分散、表述多样、逻辑依

赖隐式化和人类理解困难等问题，研究可计算的数学知识表达。面向形式化难度较高、语义结构复杂的数学对象，发展面向数学自然语言的算子树等结构化表达理论，开发相应的解析工具，支撑概念归一与语义规范化。面向多源数学资源，建立覆盖概念、定义、命题、证明和依赖关系的系统级数学知识图谱，实现分散知识的体系化组织、持续更新和关联查询。面向数学内容的可视化生成，研究数学对象、结构关系和语义约束的表达、校验与修正方法，建立可靠数学作图技术和数学家可理解、可审阅、可修正的约束表达机制。基于上述可计算数学知识表达工具，构建多模态数学语料库，研发可信数学内容审校与图文公式一体化查询系统，支持文本、公式、证明和图形等对象的统一检索、相互关联与一致性校验，服务数学文稿、数学教育和数学研究。

考核指标：形成涵盖结构化自然语言数据至少 50 万条，知识图谱节点至少 50 万个、依赖关系至少 500 万组、数学图形至少 1 千类的多模态数学语料库，支持静态资源整理、动态增量更新和专家校验闭环。使用项目研发工具所得的结构化自然语言与自 Lean 4 形式化数学复原得到的类自然语言语法树之间，通过广义树编辑距离计算的语义相似程度不低于 85%；跨表达语义相似性的准确率不低于 85%。建立面向数学知识的内容审校，支持百万级图谱在线查询，语义检索 Top-10 召回率不低于 85%，平均查询响应时间不超过 3 秒。

在固定模型、测试集和推理成本下，接入本项目知识图谱后，研究生课程与前沿科研水平的定理证明成功率分别不低于70%（对比 GPT 的成功率约为 56%）、30%（对比 LeanSearch v2 的成功率约为 20%）。建设不少于 1 万个带有“对象-关系-约束”结构化标注的数学图形样本，覆盖不少于 5 类数学方向和 20 类典型数学约束。在不少于 1000 个专家标注测试样本上，图形-语义一致性通过率不低于 85%。技术就绪度(TRL)达到 6 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：可计算数学知识表达，知识图谱，结构化自然语言，数学语料库，交互平台。

经费说明：中央财政资金 300 万元。

3.2 人工智能辅助的基础数学里程碑成果高质量形式化（青年科学家项目类）

研究内容：围绕基础数学里程碑成果，开展系统性形式化仓库建设，构建高质量形式化数学基础设施。里程碑成果建立在多领域深刻发展之上，如韦伊猜想、庞加莱猜想等，形式化这些成果将牵引大量基础理论的系统性建设，高质量产出可构成前沿数学形式化所必须的基础设施。针对选定里程碑成果，梳理证明结构与依赖关系，完成形式化任务分解与模块接口对齐。发展形式化仓库辅助或自动审核与持续维护技术，检测代码风格、冗余依赖、接口一致性等；开发适

应上游库演进的自动重构工具，确保仓库长期可维护，支持多人协作与社区贡献。第三，研究高质量形式化方法论，以人机结合方式汲取形式化专家对抽象层次与接口设计的经验，嵌入自动化流程；发展面向易读性、可维护性、可复用性的设计原则，使代码可贡献至主流形式化库或作为基础设施被下游项目依赖。构建高质量形式化数学库，为数学知识的全面可验证化提供基础设施支撑。

考核指标：选取至少 1 项里程碑定理，如韦伊猜想或庞加莱猜想，完成如德利涅-洛蒙定理，典范邻域定理的形式化，要求构建包含全部前置理论、不依赖任何未形式化前提的关键环节目标定理的完整证明链。形式化条目不少于 4000 条。建成开放的里程碑形式化数学库，形式化代码不少于 10 万行，经自动审核修改接近形式化社区最高标准，向上游仓库提交并被接受的合并请求不少于 20 个。核心形式化定理库被不少于 5 个下游形式化项目所依赖，体现其对其它证明任务的支撑作用。协作平台支持 10 人以上数学家与形式化专家协同工作，具备任务拆分与自动验证功能。面向国际学术界开源形式化仓库及维护工具集。技术就绪度（TRL）达到 6 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：数学形式化，交互式定理证明，高质量形式化，自动审核与维护。

经费说明：中央财政资金 300 万元。

3.3 面向国产自主基础模型的数学推理后训练增强方法 (青年科学家项目类)

研究内容：与国际顶尖模型相比，国产自主基础模型在复杂数学任务中的推理正确性、自我校验以及跨题型、跨学科迁移能力等方面仍存在不足。针对这一瓶颈，研究基于国产自主基础模型的数学推理后训练方法：采集自然语言策略探索、形式化尝试、验证反馈等数学推理轨迹数据，构建可学习、可复用的高质量数学推理训练数据；按照“基础推导-复杂推理-开放问题求解”的递进路径，研究由易到难、由局部到整体的课程组织与动态调度后训练方法，缓解能力失衡、知识遗忘和训练不稳定问题；基于国产开源基础模型开展数学推理后训练，使模型能够动态选择自然语言推理、形式化证明与定理检索，在国际公认的中高难度数学问题集上验证其推理正确性、求解效率、鲁棒性和迁移能力。

考核指标：构建不少于 2 万条高质量数学推理训练数据，其中成功求解数据不少于 5000 条，包含失败诊断与路径修正信息的数据不少于 1 万条，包含 30 - 50 个可验证推理单元的长程数据不少于 2000 条；数据经形式化验证、程序执行或专家抽检的正确率不低于 95%，重复率不高于 1%。提出面向国内自主基础模型的数学推理后训练与分层课程训练方法，完成不少于 2 个具有代表性的国产开源自主基础模型适配。在不少于 5 个国际公开数学推理基准上开展评测，其中至少包

含 2 个自然语言数学推理基准、2 个形式化数学基准。在保持基础模型、推理预算、工具权限和评测设置一致的条件下，与未经专项数学后训练的同源基础模型相比，模型在各基准上的平均正确率或形式化证明成功率绝对提升不低于 10 个百分点，在长程推理任务子集上的相对提升不低于 20%。在题型变化、学科分布变化、推理难度提升或表达方式变化等场景下，模型相对于分布内测试的性能下降幅度较原始基础模型减少 25% 以上。研发面向国产开源自主基础模型的数学推理后训练模型原型，开源训练代码、评测工具和数据。技术就绪度（TRL）达到 6 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：国产自主基础模型，数学推理模型，课程化训练，后训练方法。

经费说明：中央财政资金 300 万元。

3.4 程序的功能正确性与安全属性验证（青年科学家项目类）

研究内容：研究面向程序验证场景的神经符号验证框架，探索神经模型辅助规约生成、不变量与引理构造及证明搜索的方法，结合符号验证反馈修正候选结果，确保核心验证过程基于形式数学、验证结论经形式系统核验，避免依赖大模型的直觉判断。面向实际 C/C++ 程序，研究程序形式语义、验证条件生成与模块化证明方法，重点提升对指针操作、动

态内存和并发共享状态的验证能力，开发支持规约构建、形式化证明和反例分析的程序验证工具。围绕常见算法与数据结构，完成重要数学结论的形式化表达与证明，研究自然语言程序需求向形式化功能规约的转换及需求与规约的一致性检查，建立算法数学性质、数据结构抽象与程序实现之间的形式化关联，形成可复用的规约、引理和证明组件。基于上述工具与组件，开展操作系统内核、编译器、密码库、网络协议栈等基础软件源代码的内存安全性、并发安全性和功能正确性验证，建设源代码、形式化规约、验证条件、证明证据和验证结果相互关联的代码库，支持验证过程的追溯与复核。

考核指标：形成 1 套基于神经符号机制的程序验证工具原型，核心证明过程由基于逻辑的符号化内核检查，支持指针、数组、结构体、联合体、嵌套数据结构、动态内存、循环、递归及链表、树、图等典型数据结构，能够验证内存安全、算术安全与功能正确性。构建算法与数据结构相关程序验证数据集，覆盖不少于 2 门本科生程序设计或算法类课程的主要内容，包含 C 函数不少于 1000 个、C 代码不少于 20 万行，覆盖同一算法或数据结构的多种实现方式；其中包含自然语言需求、形式化规约、辅助断言、验证结果及错误诊断等信息的高质量样本不少于 5000 条。针对本科算法与数据结构课程中的标准程序的 C 语言源代码，在不得修改程序主

体的条件下，自动生成形式化规约和辅助断言并完成验证的成功率不低于 90%，该数据集应当包含指针、动态内存和复杂数据结构的程序子集上。所有通过验证的结论均能够生成可由基于逻辑的符号化内核检查。技术就绪度（TRL）达到 6 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键字：神经符号方法，形式化验证。

经费说明：中央财政资金 300 万元。

3.5 面向同调代数、层论与 Topos 理论的 AI 驱动逻辑环境与数学发现（常规项目类）

研究内容：面向同调代数、层论与 Topos 理论中数学对象高度抽象、跨理论语境迁移复杂、概念发现与理论组织困难等问题，研究 AI 驱动的逻辑环境与数学发现方法，研发具备结论迁移判断、关键结构识别、反例构造、概念生成与理论修正能力的数学研究智能体。构建数学对象、结构条件与变换操作的统一形式化表示及可追溯理论演化图谱，建立数学文献、自然语言论证与形式化证明系统之间的双向接口；研究结构迁移、假设变化与反例反馈驱动的问题提出、猜想生成和概念提炼方法，开展一致性、新颖性与数学价值评估；探索核心概念识别、理论单元组织与研究方向发现方法，支持猜想族和研究纲领的持续演化；集成文献分析、结构变换、猜想生成、反例搜索、证明规划与形式化核验等多智能体，

构建交互式数学研究环境，围绕同调不变量、导出函子、下降理论及 Topos 内部逻辑开展验证，实现研究过程可追溯、结果可复查、理论结构可演化。

考核指标：完成面向同调代数、层论与 Topos 理论的形式化逻辑环境，形成不少于 50,000 行形式化代码；围绕不少于 8 个代表性数学任务开展端到端验证，覆盖上述三个方向，其中不少于 6 项形成可由 Lean 等证明系统检查的形式化结果；形成不少于 6 类可复用论证单元，每类至少在 2 个独立任务中完成验证。研发结构与条件分析、猜想变换、反例构造、证明组织和形式化核验等专用智能体及协同机制，在不少于 20 个独立典型任务上开展测试，其中不少于 12 项完成具有明确数学意义的分析、修正、证明规划或反例构造，不少于 8 项完成从问题输入到证明或反例核验的完整流程。完成候选猜想提出、筛选、反例检验或证明核验，形成不少于 2 项具有代表性的严格数学结果，并至少覆盖反例或必要条件识别、命题修正并得到成立结论、完整证明或关键步骤形式化核验三类中的两类。技术就绪度达到 TRL 6 级以上，代表性论文中发表在我国科技期刊上的比例不低于 20%。

关键词：数学问题与猜想发现，数学理论体系构建，反例构造，形式化逻辑环境，数学理论图谱，研究型数学智能体，同调代数，层论，Topos 理论。

经费说明：中央财政资金 1000 万元，项目配套资金 500

万元。

3.6 面向计算数学的算法智能发现与可信验证（常规项目类）

研究内容：针对人工智能开展计算数学基础算法研究时存在算法知识分散、长链条推理和算法构造能力不足、理论结论与计算结果难以协同验证，以及算法失效条件和适用边界难以自动识别等问题，研究面向计算数学基础算法与新型算法发现的智能体逻辑环境与可信方法。构建结构化知识体系及问题、算法、反例与失败案例库，研究算法、理论性质、适用条件与验证证据的形式化或半形式化表示及可追溯关联，支撑智能体训练、推理与评价。研究多智能体协同、专业工具调用及验证反馈驱动的推理方法，融合形式证明、符号推理、数值验证和程序验证，开展反例搜索与构造，识别算法失效条件与适用边界，驱动算法候选和理论结论修正。围绕数值代数、反问题计算和量子计算等方向，研究知识与数学结构协同驱动的算法检索、生成、组合与优化方法，构建贯通问题理解、算法构造、理论分析、程序实现与实验验证的智能体系统，形成具有明确数学依据的新算法、重要算法改进及相关理论结果。

考核指标：建成 1 套面向计算数学基础算法与新型算法研究的智能体逻辑与知识环境，形成不少于 40000 条形式化或半形式化知识单元，其中不少于 20000 条完成形式化表示

并通过语法与类型检查；建立不少于 100 个标准问题与基础算法案例，完成不少于 50 项算法理论性质或实现可靠性的机器辅助验证，其中不少于 20 项关键性质通过形式证明系统内核检查；形成并验证不少于 10 项数值代数、反问题计算和量子计算等方向的新算法，经人类专家确认具有正确性、优越性和可复现性。构建多智能体协作系统。技术就绪度（TRL）达到 5 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：基础算法发现，智能体逻辑环境，可信验证，数学推理。

经费说明：中央财政资金 1000 万元，项目配套资金 500 万元。

3.7 基于符号学习的物理学规律合情推理形式化与多层验证体系（常规项目类）

研究内容：面向物理学研究中推理过程难以结构化记录、假设条件与近似边界不清、符号关系缺乏物理解释、验证标准不统一、推理结论置信程度难以量化等问题，研究物理合情推理的形式语义、类型化表达与多层验证方法。以物理学中反复出现的精确推导、受控近似、非受控近似、对称性论证、量纲与标度分析、极限分析、自洽性论证和数值归纳等推理模式为对象，建立统一的结构化表示、物理推理语言与逻辑依赖模型，显式描述物理量、假设、近似、尺度、边界

条件、证据来源及结论强度与适用范围，研究大语言模型驱动的物理文本、公式推导和数值结果向可检查推理链的自动转换方法。研究物理先验约束下的符号学习与假设筛选方法，将候选符号关系作为可检验假设纳入合情推理链，形成候选生成、条件检查、证据关联与推理修正的迭代机制。协同符号计算、数值仿真与形式验证，针对精确推导、受控近似、非受控近似和数值归纳，建立涵盖严格证明、误差分析、稳健性检验和数值交叉核验的分层验证体系，研究考虑证据强度、验证层级与推理依赖的置信度量化及传播方法。围绕统计物理、高能物理、凝聚态物理等真实科研问题开展应用验证与科研效能评估，形成可记录、可检查、可追溯、可比较的物理合情推理闭环，支撑可信物理规律发现与格物智能逻辑环境建设。

考核指标：形成科学合情推理形式化与验证系统，覆盖不少于 8 类典型科学推理模式和 7 类自动或半自动校验机制；完成不少于 10 万条推理链结构化转换，其中不少于 1000 条研究级完整推理链经专家审校，关键要素完整表达率不低于 90%。教材级推导链验证成功率不低于 80%，研究级推导链关键步骤验证覆盖率不低于 60%，错误检出率不低于 90%、误报率不高于 10%。围绕不少于 30 项真实科研任务、至少 3 家独立科研团队开展应用验证，推理链整理与核验人时降低不少于 50%，关键验证结论独立复现率不低于 90%。在不少

于3个自然科学方向完成不少于10项研究级端到端验证案例，每个方向不少于3项；对不少于1000个关键近似步骤开展定量置信评估，置信度校准误差不高于10%，与领域专家判断一致率不低于85%。技术就绪度（TRL）达到6级以上，研究形成的代表性论文发表在我国科技期刊上的占20%以上。

关键词：科学合情推理，形式化表达，多层验证，符号学习，物理规律研究。

经费说明：中央财政资金1200万元。

3.8 基础科学逻辑知识库与知识驱动发现研究（常规项目类）

研究内容：面向理论物理知识组织与科学发现，建设可与其他自然科学领域扩展的基础科学逻辑知识库。以“假设-推导-近似-结论-验证”为基本单元，研究多源知识抽取与对齐、逻辑依赖组织、版本与证据管理、多难度推理评测及受控知识回写，形成显式记录逻辑关系、适用条件与验证依据的可计算、可追溯知识网络。利用网络结构、逻辑缺口、矛盾关系及跨领域结构相似性识别科学问题，在已有知识、科学规律与适用条件约束下生成候选假说，结合符号推导、数值计算、文献与实验证据开展分层验证，将经验证的新结果纳入知识库。以凝聚态物理与功能材料为重点，围绕材料结构、物性、机制、计算方法、误差与实验证据的逻辑关联，研究第一性原理方法组织与调用、候选材料生成筛选、逆向

设计及反例搜索，在交错磁性、高温超导、拓扑材料和钙钛矿功能材料等体系中验证；面向统计物理、量子多体、量子信息、高能物理及非平衡过程开展系统验证，检验通用表示、评测与工具接口向化学、生物等领域的扩展能力。

考核指标：建成基础科学逻辑知识库 1 个，经验证科学命题规模不少于 500 万条，覆盖凝聚态物理、量子信息等理论物理领域，以及其他不少于 5 门自然科学核心学科的推理骨架。建成博士生级和前沿研究级跨学科科学推理冻结评测集，每学科每级不少于 100 项，总量不少于 1000 项，任务具有明确问题定义、可核验判据和来源标注。统一测试条件，测评不少于 3 个国内外代表性顶尖模型，前沿研究级最优模型任务通过率不高于 30%。基于逻辑知识库，在凝聚态物理、量子信息、非平衡统计物理、第一性原理方法等理论物理前沿方向产出不少于 5 项经独立验证并发表在物理顶级期刊的原创科学成果。在至少 2 类不同学科之间发现可验证的跨领域推理结构迁移线索，并形成经专家认可的科学问题或候选假说。技术就绪度（TRL）达到 6 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：基础科学逻辑知识库，自然科学知识体系，知识驱动发现，理论物理，科学知识网络，科学问题生成，分层验证，第一性原理计算，功能材料。

经费说明：中央财政资金 1100 万元，项目配套资金 700

万元。

3.9 多源约束与领域规律驱动的基础科学计算推理与可执行验证方法研究（青年科学家项目）

研究内容：针对基础科学问题难以直接转化为可执行计算任务、生成代码缺少领域规律约束、数值结果缺乏系统校验与可复核证据等问题，研究多源约束与领域规律驱动的基础科学计算推理与可执行验证方法。研究自然语言、公式、数据和已有程序的解析对齐与计算任务表示，显式描述问题假设、控制方程、近似条件、参数、初边值条件、适用范围和验收判据，形成规范化任务与冻结评测集。将量纲、守恒、对称性和极限行为等领域规律转化为代码生成约束、运行时检查与自动测试，研究数值方法与求解器选择、代码生成与执行及验证反馈驱动的自纠正方法，建立可复现的计算执行链。研究收敛性、稳定性、误差、残差、敏感性与不确定性等多层校验方法，建立跨方法、跨求解器及独立程序实现的交叉复核机制，区分数值误差、模型误差、数据误差与假设失效，形成涵盖问题定义、代码环境、计算结果、误差、适用边界和复核记录的标准化证据包与可核验结论。整合知识、数据、代码和计算资源，构建贯通“问题形式化-代码生成与执行-数值校验-证据回写”的原型系统，在理论物理等代表性科研场景中验证复杂科学问题的计算求解、规律校验与独立复核能力。

考核指标：建立基础科学计算推理与可执行验证方法体系，研制不少于 1 套贯通科学问题形式化、代码生成与执行、数值校验、结论形成和证据回写的原型系统。建设覆盖不少于 20 类计算科学任务的冻结评测集，包含不少于 1 万个可执行计算验证任务；每项具有明确的问题定义、代码或求解配置、参考结果或核验判据、误差容限和适用边界，经参考程序复算、自动检查和领域专家审核，题目有效率不低于 95%。在充足且统一的计算预算、通用代码执行环境和必要基础工具条件下，对不少于 3 个国内外代表性前沿大模型测评，表现最优模型任务成功率不高于 60%。项目系统在同一冻结留出评测集上的任务成功率较表现最优的前沿大模型提高不少于 20 个百分点。形成标准化计算证据包和复现机制；面向理论物理等领域选择不少于 2 个方向，完成不少于 100 项具有实际科研背景的端到端计算验证案例。技术就绪度（TRL）达到 6 级以上，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：计算科学，科学代码生成，可执行验证，数值方法，领域规律，计算证据，理论物理，知识驱动科学发现。

经费说明：中央财政资金 300 万元。

3.10 自增长逻辑环境与多智能体协同演化（常规项目类）

研究内容：以统计物理、量子物理等理论物理方向为核心，研制可独立运行的自增长逻辑环境与多智能体协同演化

系统，研究探索、验证、批评和裁决等多角色科研智能体的协同机制，构建“生成-验证-反驳-修正-沉淀-再研究”的持续递归演化闭环。基于系统运行数据，将知识规模、证据质量、任务难度和验证反馈表征为可观测信息场或有效信息势，建立角色分化、协作竞争、知识回写与能力增长的粗粒化动力学模型，研究环境规模、智能体数量、交互拓扑和验证强度对系统增长、饱和、相变、锁死与失稳的影响，形成面向稳定运行与科研效能提升的角色配置、交互机制和验证强度优化方法。在理论物理自动化科学发现任务中验证系统的稳定性与有效性，并逐步开展向其他自然科学领域的迁移验证。

考核指标：建成可独立部署运行的自增长逻辑环境与多智能体协同演化系统，交付代码、接口、验证工具及文档；在真实科研任务驱动下连续稳定运行不少于 180 天，完成不少于 1 万个可追溯、可验证的有效研究闭环。较冻结初始环境，经验证的有效环境规模增长不低于 100%，新增内容第三方抽检合格率不低于 90%。设置不少于 6 个环境规模或多智能体配置档位，对照实验规模覆盖不少于两个数量级；建立不少于 2 类协同演化粗粒化或统计物理模型，经真实日志标定与留出检验，复现不少于 2 类演化现象，形成定量关系与优化规则；模型和测试任务一致时，复杂理论物理任务有效解决率较冻结基线相对提升不低于 30%。覆盖不少于 4 个前沿自然科学方向，其中理论物理不少于 3 个，完成跨领域迁

移验证；形成不少于 100 项可独立构成论文核心结论的原创成果。TRL 达到 5 级以上，代表性论文在我国科技期刊发表比例不低于 20%。

关键词：自增长逻辑环境，多智能体协同演化，递归自演化，理论物理，自动化科学发现，非平衡统计物理，信息势，粗粒化动力学，尺度规律。

经费说明：中央财政资金 850 万元，项目配套资金 400 万元。

4. 认知功能的基本原理探索与原型系统构建

4.1 思考行为的数学建模与多任务认知过程解析（青年科学家项目，拟支持 3 项）

研究内容：以理解“思考”现象为核心，针对“思考”这一复杂认知行为开展系统性的数学建模研究，构建具备普适性的思考理论。在此基础上，深入剖析并解释多种典型的思考方式与策略，包括但不限于快慢思考、搜索机制、问题拆解、思维链等，揭示其在不同认知场景下的运作机理。进一步地，将理论应用于具体复杂任务的求解过程，阐释在典型场景中，有效解决问题的具体思考路径与策略组合，从而实现从抽象思考方式到具体任务执行全过程的理论解释与建模闭环。场景可以包括但不限于以下情况：数学推理问题、代码生成、科学探索、电子游戏、棋牌对弈、具身操作与长程任务规划、复杂系统故障诊断与根因定位、开放域深度调

研与工具使用、工程与分子/材料/芯片设计优化、组合优化与实时资源调度、医疗诊断与临床决策等。

考核指标：提出 1 套针对思考行为的数学建模理论，该理论需具备严谨的数学形式化表达，能够统一描述思考行为的形成方式。理论需能够有效解释至少 3 种常见的思考方式或策略，如快慢思考、搜索、问题拆解、思维链等，并阐明各方式之间的内在联系与适用边界。针对至少 3 类具体任务场景，利用所建理论清晰解释其有效解决问题的具体思考过程与策略应用，做细致且具体的分析，形成可验证的认知过程解析案例。技术就绪度（TRL）指标达到 TRL-3 关键功能实验验证，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：思考数学建模，理论，思考方式，任务思考过程，问题拆解，思维链。

经费说明：每个项目中央财政资金 300 万元。

4.2 记忆行为的数学理论建模、计算原型和验证平台（常规项目）

研究内容：围绕记忆这一复杂认知行为，建立能系统解释记忆现象、揭示记忆机理并描述从记忆到知识演化过程的数学建模理论，并以此为基础研制可计算、可复现的记忆计算原型系统与开放式验证平台。（1）建立记忆的统一数学理论。该理论能够统一解释认知科学所揭示的记忆主要类型，

包括但不限于短期记忆、工作记忆、长期记忆，以及陈述性记忆、非陈述性记忆，描述记忆的编码、存储、巩固、遗忘和提取等关键过程，形成能够解析分析和计算检验的记忆核心理论，重点揭示以下规律：①多时间尺度表征的动力学：研究工作记忆、短时记忆和长时记忆在时间尺度上的差异化表征机制与演化模式，揭示不同类型记忆在持续信息更新条件下维持稳定性的内在数学规律。②表征空间的几何与拓扑结构演化：揭示短期与长期、陈述性与非陈述性记忆在表征空间中聚合、分化、迁移与稳定化的动力学法则，阐明层次化记忆组织对知识存储、检索与泛化的支撑机制。③从记忆到知识的相变理论：揭示记忆如何通过选择性巩固、压缩、重组和泛化实现去情境化并涌现为结构化知识，给出这一转化的动力学条件或相变判据。（2）构建记忆的计算原型系统。基于上述理论，构建可计算可验证的算法原型系统。该原型系统应具备多类型记忆的计算模拟能力，包括但不限于短期记忆、工作记忆、长期记忆，以及陈述性记忆、非陈述性记忆，支持对不同记忆过程的参数化操控与动力学观测，能够根据理论法则自动分化出具备不同时间尺度与抽象层次的记忆层，突破依赖人工设计记忆层级与固定存取流程的局限。

（3）构建开放式验证平台。构建集认知实验数据管理、模型仿真运行、量化指标评测、结果可视化与第三方复现于一体的开放式验证平台。平台需集成序列学习、联想记忆、情景-

语义记忆交互等典型记忆场景的标准测试基准，支持模型输出与人类行为数据及神经活动数据的多维度比对，形成标准化、自动化的模型验证平台。

考核指标：提出一套统一的记忆数学理论框架，该框架需具备严谨的数学形式化表达，明确公理化假设，覆盖短期记忆、工作记忆、长期记忆以及陈述性记忆、非陈述性记忆等至少 5 类主要记忆类型，能够形式化描述并推导记忆的编码、存储、巩固、遗忘和提取过程，能够给出从碎片化感知记忆到结构化知识涌现的动力学条件或相变判据。基于理论构建记忆计算原型系统，需支持至少 3 种记忆过程的参数化操控与动力学观测，如编码、存储、巩固、遗忘、提取等，并能在公开记忆评测基准上完成验证，在记忆保持率、遗忘率等至少 2 个核心指标上相比于当前基线方法取得更优的表现。构建记忆行为建模的开放式验证平台，需包含序列学习、联想记忆、情景-语义交互等至少 3 类经典场景，每个场景需收集不少于 10 个可验证的记忆行为认知解析过程案例和有关数据、以及不少于 2 个对比基线，就数学理论和计算模型产生的定量预测产生进行实验检验和基线对比，开源完整的数据、基线和评测协议，确保第三方可独立完整复现评测全流程。技术就绪度（TRL）指标达到 TRL-5 子系统相关环境验证，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：记忆数学建模，统一记忆理论。

经费说明：中央财政资金 2500 万元，配套资金不少于 900 万元。

4.3 记忆行为的数学理论建模（青年科学家项目，拟支持 2 项）

研究内容：围绕记忆这一复杂认知行为，建立能系统解释记忆现象、揭示记忆机理并描述从记忆到知识演化过程的数学建模理论。该理论能够统一解释认知科学所揭示的记忆主要类型，包括但不限于短期记忆、工作记忆、长期记忆，以及陈述性记忆、非陈述性记忆，描述记忆的编码、存储、巩固、遗忘和提取等关键过程，揭示记忆如何通过选择性巩固、压缩、重组、泛化实现去情境化、形成结构化知识的数学规律，形成能够解析分析和计算检验的记忆核心理论。选取序列学习、联想记忆、情景-语义记忆交互等典型场景，通过实验数据分析与计算仿真相结合的手段，检验模型对记忆行为的解释和建模能力。

考核指标：提出 1 套统一的记忆数学理论框架，该框架需具备严谨的数学形式化表达，明确公理化假设，覆盖短期记忆、工作记忆、长期记忆以及陈述性记忆、非陈述性记忆等至少 5 类主要记忆类型，能够形式化描述并推导记忆的编码、存储、巩固、遗忘和提取过程。构建从记忆到知识的转化模型，给出从碎片化感知记忆到结构化知识涌现的动力学

条件或相变判据。根据理论模型产生可被实验检验的定量预测，通过合作实验或公开数据针对序列学习、联想记忆、情景-语义记忆交互等至少 3 类任务场景进行计算仿真，形成可验证的认知过程解析案例，建立标准化的验证平台，提供完整的数据和基线，使得结果可在第三方环境中复现与评测。技术就绪度（TRL）指标达到 TRL-5 关键功能实验验证，研究形成的代表性论文发表在我国科技期刊上的占 20% 以上。

关键词：记忆数学建模，统一记忆理论。

经费说明：每个项目中央财政资金 300 万元。

4.4 统一视觉计算基础理论与智能原型系统（常规项目）

研究内容：生物视觉系统能在视角变化、遮挡、观测不全和光照变化等条件下，从二维观测中形成稳定的三维场景理解，并据此支持空间认知与行动决策。这背后蕴含着一组关于视觉感知与推理的基本问题，如对象与场景如何被稳定识别、空间关系如何从部分观测中被推断、动态变化如何被感知与预测、不确定性如何影响判断与行动等。从这些基本问题出发，探究视觉感知与推理中对象与场景识别、空间关系推断、动态变化感知、遮挡补全、行动条件判断等关键构成要素的背后原理，建立这些要素之间的数学关系与内在逻辑。在此基础上，提出一套可计算可验证的统一视觉计算理论，能够阐明上述多个视觉现象背后的基础原理，刻画视觉信息在感知、推理、生成与决策各环节中的表征演化与传递

规律。进而，探究视觉主导的智能学习新范式，使系统能够主要依赖视觉观测实现场景理解、空间推理、状态预测与决策。在上述理论与范式的基础上，构建视觉驱动智能体，并应用于三维/四维重建、三维视觉/场景理解、生成任务等多个典型视觉任务，形成从基础理论到模型与学习范式设计、再到工程应用的全链条支撑，并研制原型系统完成验证。

考核指标：提出 1 套视觉计算基础数学理论，该理论需具备严谨的数学形式化表达，能够统一刻画视觉信息从感知到决策的演化过程。该理论需能够解释至少 3 类基本视觉现象，如对象与场景如何被稳定识别、空间关系如何从部分观测中被推断、动态变化如何被感知与预测等，阐明各类现象背后的内在机制及其适用条件。针对至少 3 类典型视觉任务场景，如动态场景理解与生成、三维/四维空间感知与重建等，利用所建理论框架理解并推导出其有效解决问题的具体感知与决策过程，形成可验证的视觉认知过程解析案例。在此基础上，提出视觉主导的智能模型学习范式，构建视觉驱动智能体原型系统，在公开三维重建基准，如 ShapeNet、CO3D 等上完成验证，相比主流方法在数据和模型效率，精度，质量，场景级结构一致性等性能指标上改进不低于 10%。在室内场景、动态室外场景、物体为中心场景等至少 3 类典型场景中完成验证。技术就绪度（TRL）指标达到 TRL-5 子系统相关环境验证，研究形成的代表性论文发表在我国科技期刊

上的占 20%以上。

关键词：视觉认知，统一视觉计算理论，纯视觉智能体，空间推理，视觉理解与生成。

经费说明：中央财政资金 2500 万元，配套资金不少于 900 万元。

4.5 统一视觉计算基础理论（青年科学家项目，拟支持 2 项）

研究内容：生物视觉系统能够在视角变化、遮挡、观测不全和光照变化等条件下，从二维观测中形成较为稳定的三维场景理解，并据此支撑空间认知与行动决策。这背后涉及一系列关于视觉感知与推理的基本问题，例如对象与场景如何在变化条件下被稳定识别、空间关系如何从部分观测中被推断、动态变化如何被感知与预测、不确定性如何影响判断与行动等。围绕上述问题，建立一套可计算可验证的统一视觉计算理论，系统刻画视觉感知与推理中对象与场景识别、空间关系推断、动态变化感知、遮挡补全、行动条件判断等关键构成要素，探究其背后的基础原理，阐明视觉信息在感知、推理、生成与决策各环节中的表征演化与传递规律。在此基础上，进一步提出视觉主导的学习范式与智能体基本设计原则，并在典型视觉任务中开展理论验证。

考核指标：提出一套面向视觉智能的感知与空间推理基础理论，具备明确的数学定义、基本假设和可验证的推导关

系，能够刻画视觉观测、结构化表征、空间推理、状态预测与决策支持之间的演化过程；针对至少三类基本视觉现象，如对象与场景在视角变化、遮挡或光照变化下的稳定识别、部分观测条件下的空间关系推断、视觉信息不确定性对场景判断和状态预测的影响等，给出统一理论框架下的机制解释及适用条件；基于所建理论，进一步提出视觉主导智能的学习范式与模型基本设计原则，阐明视觉表征、状态更新、空间推理等关键环节的内在原理，为视觉智能体的构建提供可推导、可验证的理论依据；针对三维场景重建、动态场景预测、视觉感知与理解等至少三类典型视觉任务，形成由理论指导的视觉认知过程解析案例，并在公开数据集或可控仿真环境中完成验证。技术就绪度（TRL）指标达到 TRL-3 关键功能实验验证，研究形成的代表性论文发表在我国科技期刊上的占 20%以上。

关键词：视觉智能，视觉计算理论，空间推理，感知与推理。

经费说明：每个项目中央财政资金 300 万元。

4.6 规则表示与推理行为的数学建模（青年科学家项目，拟支持 2 项）

研究内容：面向当前大模型主要依赖统计模仿、难以自发形成推理行为的问题，本项目旨在探明构成推理行为的底层认知单元及其计算原理，建立能够描述归纳、演绎、溯因

等多种推理模式的现象层面的数学模型，为最终建立能够统一推导出多种认知现象的基础理论、设计能够自主发现常识、规则乃至科学定律的下一代人工智能奠定基础。主要研究内容包括：系统梳理并发现构成各类推理行为的底层认知操作，识别其基本功能单元。研究不预设具体清单，鼓励从认知科学实验、发展心理学观察以及大语言模型的行为分析中，提炼出不可再分的原子操作。可关注的认知现象包括但不限于：特征与模式的主动识别，研究系统如何在与环境交互中使共现特征得到突显，形成初步的共同特征表征；感知与动作的层级抽象，探究从具体感知或底层动作序列中形成视觉概念与宏观动作的认知过程；猜想的生成与价值判断，研究系统形成假设的机制，以及驱动系统判断猜想价值、决定是否将其提升至显意识进行验证的评估机制；案例的主动搜索与构建，揭示系统为证明或证伪猜想，如何在思维空间或外部环境中构建具体的正例反例；深入提问，递归式生成新问题以最小化对象的不确定性。针对提出的原子操作，构建其数学描述与计算模型，模型必须满足可证伪性，并能解释人类认知实验的结果或大模型的推理行为。建模工作不预设具体清单，可关注但不限于以下方向：内蕴奖励与价值函数的建模，将猜想的价值建模为某种内蕴奖励的代理，比如环境不确定性的下降；共同特征识别的建模，阐明短期记忆如何突显重复出现的特征；感知与动作粗粒化的建模，理解其涉及的时

间尺度分化的机制；案例模拟的建模，分析情景式或语义式推演的实现方式。

考核指标：识别并定义至少 5 个构成推理行为的底层认知基本单元。需具有明确的、可操作的定义，其界定需能涵盖归纳、演绎、溯因、类比等至少 2 类不同的推理行为。针对至少 3 个基本单元，提出其可能的工作原理模型，并验证建模的有效性。验证手段可包括但不限于认知神经科学实验数据的定量拟合、数值实验复现关键行为特征、数学定理证明。考核以获得新知识、新理解为核心导向。鼓励发现此前未被充分描述的认知单元，或对已知现象给出具有更强解释力与预测力的新原理。

设计并构建至少 3 个具体的任务场景，用于展示所识别的基本单元如何在交互中协同工作，驱动完整的推理行为涌现。场景设计应遵循无先验知识介入原则，即系统初始不具备目标规则或概念的显式编码等人为设计。场景选择应具有多样性和代表性，可从以下方向选取但不限于此：①工具交互与规则发现，在不具有先验知识的情况下，通过操作虚拟计算器、几何沙盘等工具，自主推理出四则运算规则或平面几何定理；②语言与符号系统的涌现，通过与其他模型或智能体沟通协作，自发形成具有组合性语法规则的语言或符号协议；③物理与社会常识的习得，通过与物理环境、模拟环境或网络信息环境的互动，推理出符合人类直觉的物理常识

或社会行为规范。技术就绪度（TRL）指标达到 TRL-3。研究形成的代表性论文发表在我国科技期刊上的占 20%以上。构建并开源一套用于测试和比对推理行为的仿真实验环境与基准数据集。

关键词：归纳与演绎推理，认知功能建模，抽象化表示。

经费说明：每个项目中央财政资金 300 万元。