

可解释、可通用的下一代人工智能方法重大研究计划 2026 年度项目指南

可解释、可通用的下一代人工智能方法重大研究计划面向人工智能发展国家重大战略需求，以人工智能的基础科学问题为核心，发展人工智能新方法体系，促进我国人工智能基础研究和人才培养，支撑我国在新一轮国际科技竞争中的主导地位。

一、科学目标

本重大研究计划面向以深度学习为代表的人工智能方法鲁棒性差、可解释性差、对数据的依赖性强等基础科学问题，挖掘机器学习的基本原理，发展可解释、可通用的下一代人工智能方法，并推动人工智能方法在科学领域的创新应用。

二、核心科学问题

本重大研究计划针对可解释、可通用的下一代人工智能方法的基础科学问题，围绕以下三个核心科学问题开展研究。

（一）深度学习的基本原理。

深入挖掘深度学习模型对超参数的依赖关系，理解深度学习背后的工作原理，建立深度学习方法的逼近理论、泛化误差分析理论和优化算法的收敛性理论。

（二）可解释、可通用的下一代人工智能方法。

通过规则与学习结合的方式，建立高精度、可解释、可通用且不依赖大量标注数据的人工智能新方法。开发下一代人工智能方法需要的数据库和模型训练平台，完善下一代人工智能方法驱动的基础设施。

(三) 面向科学领域的下一代人工智能方法的应用。

发展新物理模型和算法，建设开源科学数据库、知识库、物理模型库和算法库，推动人工智能新方法在解决科学领域复杂问题上的示范性应用。

三、2026 年度资助研究方向

(一) 重点支持项目。

围绕核心科学问题，以总体科学目标为牵引，拟以重点支持项目的方式资助前期研究成果积累较好、对总体科学目标在理论和关键技术能发挥推动作用、具备产学研用基础的申请项目，研究方向如下：

1. 几何对称性先验嵌入的深度神经网络。

针对现有前沿深度神经网络架构难以有效嵌入几何对称性先验的瓶颈问题，发展能够与数据对称性相适配的深度神经网络泛化理论，揭示数据几何对称性与模型等变结构之间的内在关联，构建能够捕捉数据对称性的新型深度神经网络基础模块体系。完成对不少于 3 种前沿深度神经网络基础模块（如 Transformer、Mamba、INR 等）的对称性先验改造，使其具备旋转、镜像、仿射等几何对称性的建模能力；在不少于 3 类实际应用场景中验证所提出的新型深度神经网络架构的有效性，确保在参数规模相当的条件下，性能显著优于原始方法。

2. 智能体驱动的多模态全流程数据治理平台。

构建由智能体驱动的自动化数据治理平台，支持从数据获取、治理、训练到应用的全流程：（1）构建可灵活部署、支持自动调度的算力管理底座，支持不同规模的单/多模态人工智能模型（如 FastText、SAM、YOLO、Qwen-VL 等）的混合推理与流水线编排；（2）研发基于智能体的统一数

据工程平台，覆盖多源异构数据（不少于 5 种模态）的获取与治理（支持不少于 200 个算子，流水线编排成功率不低于 97%），支持大模型训练过程中数据的动态选择和配比、以及多垂域（Code、GUI、Web 等不少于 5 个领域）智能体数据的自动合成与质量管控；（3）开发面向复杂科学发现的垂域智能体，在 Lean 数学形式化推理、蛋白质与多糖结构分析、学科竞赛策略推演等不少于 3 种典型科学任务上验证有效性。

3. 高维复杂物理约束下的生成建模与科学计算。

针对数据驱动的高维科学计算生成建模中物理一致性缺失、复杂约束难以严格满足的难题，通过引入物理守恒律与结构先验，建立严谨的生成演化理论框架；设计能够保持物理不变量的数值算法，发展流形约束下的动力学控制机制，使生成路径严格满足守恒、边界与几何等约束要求；进一步面向功能材料设计、复杂物理场反演与参数识别等典型任务，研究“模型—算法—验证”一体化技术，提升约束满足精度与结果可信度。

4. 融合物理与人工智能的极端尺度科学成像。

发展面向极端尺度的科学成像新方法，推进多尺度、多物理场计算物理模型与生成式人工智能、视觉—语言基础模型等前沿人工智能技术深度融合，构建可解释、可通用的成像反演框架。面向先进制程半导体三维量测、非侵入深层生物组织成像、深空天文探测等极端空间、时间与能量尺度场景，突破现有技术在成像分辨率、成像深度、成像对比度、成像通量及不确定性量化等方面的瓶颈，构建可开放共享的模型、数据与工具体系，在不少于 3 种典型科学成像任务上验证有效性。

5. 基于深度学习的量子多体计算。

面向多电子薛定谔方程的高精度与高可扩展求解，发展以变分优化为核心的量子多体计算方法。围绕量子多体基态、激发态、有限温度与动力学等科学问题，构建融合神经网络量子态、张量网络与生成模型的新型计算框架，突破高精度、大规模与长时间演化的计算瓶颈，建立可复用的开源算例库与数据集。在强关联体系、高压致密物质及超快动力学等前沿领域，取得可实验验证的热力学状态方程、相图与谱学计算结果。

6. 人工智能加速的仿星器优化设计。

针对仿星器优化中多目标、高参数维度与强非线性等挑战，发展人工智能驱动的仿星器聚变堆设计新方法：（1）构建开源的大规模先进仿星器平衡态与线圈设计数据库，规模不少于 10 万组，覆盖典型设计空间，支撑模型训练标定与可复现评测；（2）基于生成式人工智能技术，对指定设计指标快速生成可行的初始候选方案，实现仿星器优化设计的“热启动”；（3）建立等离子体约束与稳定性等多目标优化的高保真深度神经网络代理模型，提升单次评估效率与可信度，支撑闭环迭代优化。

7. 耐受极端环境的生物元件设计平台。

面向生物制造对耐受极端工业环境生物元件的需求，开发通用的、人工智能驱动的设计平台。（1）建立来自真实世界的大规模生物元件序列数据集，规模不少于 100 亿条，包括不少于 50 亿条环境标签，包含不少于 10 亿条来自高温、高压、强酸、强碱等极端环境的序列；（2）构建生物元件预训练人工智能模型，并开发极端环境适应度预测模型与特定条件下的元件挖掘模型；（3）研究小样本和主动学习方法，开发配套自动化实验

装置，实现干湿闭环迭代。在小样本（不超过 100 个实验样本）条件下，完成不少于 3 款可耐受极端工业环境的生物元件的设计验证。

（二）集成项目。

本年度拟遴选具有重大应用价值和良好研究基础的研究方向进行集成资助，研究方向如下：

1. 深度神经网络动力学机制理论与新架构设计。

针对深度神经网络在训练中的重要现象，从动力学角度研究现象背后的机制，为深度神经网络训练优化与新架构设计提供理论指导。（1）针对大模型训练中的尺度定律（scaling law）现象，通过系统分析学习率、批大小、权重衰减等超参数作用，揭示尺度定律产生的内在机理及边界条件，形成面向大模型训练的调参指导原则，并在参数规模不少于 10 亿的模型训练任务中验证；（2）针对深度神经网络的训练过程，通过系统分析架构、优化算法、初始化、权重衰减等，研究嵌入结构的形成机制、记忆和推理的偏好机制、不稳定现象（如损失尖峰等）的成因，以及损失景观上训练轨迹的特征等，在参数规模不少于 10 亿的模型上验证；（3）针对不同深度神经网络优化算法收敛速度存在差异的现象，选取至少三种主流优化算法，对其训练动力学特性进行分析建模，并揭示其收敛特性与内在机理；（4）基于理论研究，发展新型深度神经网络基础架构及相应预训练方法，在不少于 300 亿词元的真实数据集上进行训练，验证其相对于经典 decoder-only Transformer 架构的性能优势。

2. 符合物理规律的世界模型理论与关键技术。

面向通用智能体长期自主交互需求，研究数据驱动与物理机理融合的可交互视频世界模型技术，建立“生成—理解”一体化机制，支撑鲁棒规划、安全决策与持续学习，构建长时交互数据与物理一致性基准。（1）构建刻画连续动力学、离散事件与高层语义的统一架构，研究开放环境下模型自诊断、自完善方法；（2）提出碰撞、移动、流动等关键物理属性的显式嵌入与可控推演技术，开发支持多主体、多视角复杂场景的长时连续交互生成模型，形成百亿级参数世界模型底座，连续生成与推演时长不少于 180 秒；（3）研究物理先验驱动的因果建模与可干预推断，面向罕见事件与分布漂移实现策略演进与模型自适应，确保能量守恒、碰撞等关键物理约束违背率不超过 5%，提升对长尾风险的鲁棒性与可恢复性；（4）探索基于世界模型的不确定性评估、风险约束规划与安全决策机制，构建“世界模型+具身智能体”闭环系统原型，在灵巧操作、具身导航、多智能体协同等不少于 3 类任务中开展端到端验证，成功率不低于 85%。

3. 感知与决策协同的全病种多模态病理大模型与应用。

针对多中心病理数据分布和模态多样性的挑战，研究面向全病种、赋能临床全流程的通用病理基础大模型。（1）构建覆盖不少于 100 个中心、50 万患者的大规模病理数据库，训练不少于 300 亿参数、支持万级上下文理解的基础大模型。探索数据规模、模态配比与模型容量约束下的表征与泛化规律，研究面向病理全场图的超长上下文编码方法；（2）研究病理多模态数据（HE、免疫组化、分子病理、诊断报告）的跨模态统一表征与对齐融合机制，阐明“对齐—融合”协同驱动的跨模态语义一致性机制，构建百万级病理图文指令数据；（3）研究面向镜下阅片与高通量扫描等不

同临床场景的通用模型适配方法，发展模型量化及多规格适配的关键技术，构建面向嵌入式环境的高效推理机制，结合诊断智能体与强化学习，探索切片扫描过程中感知与决策并行耦合的主动诊断框架。在不少于 100 家医疗机构完成研究成果验证。

4. 面向空天发动机高端工程装备研发的智能基座与应用。

研究面向空天发动机高端工程装备研发的全流程智能一体化方法，并开展示范应用：（1）研究融合物理机理的大模型自监督预训练方法，构建融合多源领域知识、工程数据以及计算工具的垂域基座模型（不少于 100 亿参数），验证模型在复杂工程任务中的推理、工具调用和决策能力；（2）构建“设计—分析—制造”全流程智能体体系，实现几何设计智能生成与强泛化预训练模型的耦合，完成自动化闭环智能设计流程；（3）打通从智能生成到物理样机的全数据链路，研制不少于 3 种不同推力等级的高性能发动机样机，在真实工程型号中开展中试，验证人工智能方法在高端装备研制中的通用性与有效性。

四、项目遴选的基本原则。

（一）紧密围绕核心科学问题，鼓励基础性和交叉性的前沿探索，优先支持原创性研究。

（二）优先支持面向发展下一代人工智能新方法或能推动人工智能新方法在科学领域应用的研究项目。

（三）重点支持项目和集成项目应具有良好的研究基础和前期积累，对总体科学目标有直接贡献并发挥支撑作用。

五、2026 年度资助计划

拟资助重点支持项目约 5 项，直接费用资助强度约为 200 万元/项，资助期限为 3 年，重点支持项目申请书中研究期限应填写“2027 年 1 月 1 日 - 2029 年 12 月 31 日”；拟资助集成项目约 4 项，直接费用资助强度为 300 - 500 万元/项，资助期限为 3 年，集成项目申请书中研究期限应填写“2027 年 1 月 1 日 - 2029 年 12 月 31 日”。

六、申请要求及注意事项

（一）申请条件。

本重大研究计划项目申请人应当具备以下条件：

1. 具有承担基础研究课题的经历；
2. 具有高级专业技术职务（职称）。

在站博士后研究人员、正在攻读研究生学位以及无工作单位或者所在单位不是依托单位的科学技术人员不得作为申请人进行申请。

（二）限项申请规定。

执行《2026 年度国家自然科学基金项目指南》“申请规定”中限项申请规定的相关要求。

（三）申请注意事项。

申请人和依托单位应当认真阅读并执行本项目指南、《2026 年度国家自然科学基金项目指南》和《关于 2026 年度国家自然科学基金项目申请与结题等有关事项的通告》中相关要求。

1. 本重大研究计划项目实行无纸化申请。申请书提交日期为 2026 年 3 月 1 日 - 2026 年 3 月 20 日 16 时。

(1) 申请人应当按照科学基金网络信息系统(以下简称“信息系统”)中重大研究计划项目的填报说明与撰写提纲要求在线填写和提交电子申请书及附件材料。

(2) 本重大研究计划旨在紧密围绕核心科学问题,对多学科相关研究进行战略性的方向引导和优势整合,成为一个项目集群。申请人应根据本重大研究计划拟解决的核心科学问题和项目指南公布的拟资助研究方向,自行拟定项目名称、科学目标、研究内容、技术路线和相应的研究经费等。

(3) 申请书中的资助类别选择“重大研究计划”,亚类说明选择“重点支持项目”或“集成项目”,附注说明选择“**可解释、可通用的下一代人工智能方法**”,受理代码选择 T01,根据申请的具体研究内容选择不超过 5 个申请代码。

重点支持项目的合作研究单位不得超过 2 个,集成项目合作研究单位不得超过 4 个。集成项目主要参与者必须是项目的实际贡献者,合计人数不超过 9 人。

(4) 申请人在申请书起始部分应明确说明申请符合本项目指南中的资助研究方向,以及对解决本重大研究计划核心科学问题、实现本重大研究计划科学目标的贡献。

如果申请人已经承担与本重大研究计划相关的其他科技计划项目,应当在申请书正文的“研究基础与工作条件”部分论述申请项目与其他相关项目的区别与联系。

2. 依托单位应当按照要求完成依托单位承诺、组织申请以及审核申请材料等工作。在 2026 年 3 月 20 日 16 时前通过信息系统逐项确认提交本

单位电子申请书及附件材料，并于 3 月 21 日 16 时前在线提交本单位项目申请清单。**未按时提交项目清单的申请将不予接收。**

3. 其他注意事项。

(1) 为实现重大研究计划总体科学目标和多学科集成，获得资助的项目负责人应当承诺遵守相关数据和资料管理与共享的规定，项目执行过程中应关注与本重大研究计划其他项目之间的相互支撑关系。

(2) 为加强项目的学术交流，促进项目群的形成和多学科交叉与集成，本重大研究计划将每年举办 1 次资助项目的年度学术交流会，并将不定期地组织相关领域的学术研讨会。获资助项目负责人有义务参加本重大研究计划指导专家组和管理工作组所组织的上述学术交流活动，并认真开展学术交流。

(四) 咨询方式。

国家自然科学基金委员会交叉科学部交叉科学一处

联系电话：010-62328382