可解释、可通用的下一代人工智能方法重大研究计划2024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、2024年度资助研究方向

　　（一）培育项目。

　　围绕上述科学问题，以总体科学目标为牵引，对于探索性强、选题新颖、前期研究基础较好的申请项目，将以培育项目的方式予以资助，研究方向如下：

　　1.神经网络的新架构和新的预训练或自监督学习方法。

　　针对图像、视频、图、流场等数据，发展更高效的神经网络新架构，预训练或自监督学习方法，并在真实数据集上进行验证。

　　2.深度学习的表示理论和泛化理论。

　　研究卷积神经网络（以及其它带对称性的网络）、图神经网络、循环神经网络、低精度神经网络、动态神经网络、生成扩散模型等模型的泛化误差分析理论、鲁棒性和稳定性理论，并在实际数据集上进行验证；研究无监督表示学习、预训练-微调范式等方法的理论基础，发展新的泛化分析方法，指导深度学习模型和算法设计。

　　3.深度学习训练算法的理论基础。

　　研究神经网络损失景观的结构和训练算法的特性，包括但不限于：临界点的分布及其嵌入结构、极小点的连通性，稳定性边缘（edge of stability）和损失尖峰（loss spike）现象、算法的隐式正则化、稳定性和收敛性；训练过程对于超参的依赖性问题、神经网络记忆灾难、训练时间复杂度分析等问题；发展卷积网络、Transformer网络、扩散模型、混合专家模型等模型的收敛速度更快、时间复杂度更低的训练方法。

　　4.大模型的基础问题。

　　研究多任务、多数据、大模型的基础问题，包括但不限于大模型的表示理论和泛化理论、大模型训练的稳定性、标度率(scaling law)、涌现等现象；研究新型 (structured) state model的基础性质，包括它是否有记忆灾难(curse of memory)的困难；理解Transformer模型的表达和泛化能力、上下文学习（In-Context learning），Chain of Thoughts推理的有效性，以及模型的外推能力（例如length generalization）等。

　　5.微分方程与机器学习。

　　研究求解微分方程正反问题及解算子逼近的概率机器学习方法；基于生成式扩散概率模型的物理场生成、模拟与补全框架；基于微分方程设计新的机器学习模型和网络结构，加速模型的推理、分析神经网络的训练过程。

　　6.图神经网络的新方法。

　　利用随机游走、多项式近似、调和分析、粒子方程等数学理论解决深度图神经网络过度光滑、过度挤压、适用异配图与动态图等问题；针对药物设计、推荐系统、多智能体网络协同控制等重要应用场景设计有效的、可扩展的、具有可解释性的图表示学习方法。

　　7.人工智能的安全性问题。

　　针对主流机器学习问题，发展隐私保护协同训练和预测方法；发展面向对抗样本、数据投毒、后门等分析、攻击、防御和修复方法；研究机器学习框架对模型干扰、破坏和控制的方法；发展可控精度的隐私计算方法，数据和模型（包括大模型）的公平性、可靠性的评估与评级方法。

　　8.科学计算领域的人工智能方法。

　　针对电子多体问题，建立薛定谔方程数值计算、第一性原理计算、自由能计算、粗粒化分子动力学等的人工智能方法，探索人工智能方法在电池、电催化、合金、光伏等体系研究中的应用。

　　针对典型的物理、化学、材料、生物、燃烧等领域的跨尺度问题和动力学问题，发展物理模型与人工智能的融合方法，探索复杂体系变量隐含物理关系的挖掘方法和构效关系的数学表达，建立具有通用性的跨尺度人工智能辅助计算理论和方法，解决典型复杂多尺度计算问题。

　　9.以数据为中心的机器学习。

　　针对数据质量、数量和效率等因素，发展机器学习方法为下游机器学习模型提供大量高质量数据；针对AI for Science数据侧，研究和设计高效的科学数据（如对蛋白质和药物构图）构建和预处理方式；针对大模型数据侧，从数据获取成本和效率出发，建立科学和系统的数据质量评估策略，设计高效的数据选择方法，构建有效的数据配比方式，并探索大模型辅助数据质量提升的方法（如自动数据标注）。

　　10.基于量子计算的机器学习算法。

　　研究不同类型的学习方法如何映射到一般的量子过程，提出新算法利用量子特性实现高效学习；研究量子机器学习相对于经典机器学习方法在表达能力以及泛化能力上的优势，探索量子机器学习的可解释性，建立量子机器学习在量子物理和化学的应用场景。

　　11.开放型项目。

　　与可解释、可通用的下一代人工智能方法和AI for Science 领域相关的方法，重点支持在算法和模型方面的创新课题。

　　（二）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上能发挥推动作用、具备产学研用基础的申请项目，研究方向如下：

　　1.下一代人工智能方法。

　　发展结合逻辑推理、知识和规则的人工智能方法，建立具有可解释和可通用性的人工智能理论框架；发展适用于连续、密集数据（如图像）和非结构化数据（如分子结构）的新型神经网络架构，有效捕捉空间、结构、语义等多维度的上下文信息，提高对数据的建模能力。在真实数据集上进行验证。

　　2.新一代脑启发的人工智能模型与有效训练算法。

　　针对大脑神经元的物理形态和生物物理的多样性，建立生物神经元与人工神经元之间的简洁且有效的映射关系，使人工神经元具有生物神经元的树突非线性整合与计算功能，为建立其它类型生物神经元与人工神经元的映射提供统一的理论和算法框架。结合大脑神经元网络连接结构、脑区异质性和宏观梯度等特点，设计带生物神经元特性约束的人工神经网络模型，实现记忆、决策等高级认知功能。实现不少于3种生物神经元与人工神经元之间的有效映射和3种重要的树突计算功能，与现有映射相比，实现精度、性能和参数可解释性上的提升。

　　3.多智能体协作学习理论与方法。

　　针对多智能体协作时分布式数据处理所面对的挑战，如泛化性能缺乏保障、适应性和可扩展性弱等难题，研究高效的多智能体协作学习理论与方法，具体包括：（1）研究提升多智能体协作学习系统泛化性能的算法，分析泛化误差界；（2）针对动态变化的环境和不断扩展的网络规模，研究多智能体系统的适应性和可扩展性，确保智能体能够有效学习并高效协作；（3）在多智能体系统中处理和融合多模态数据（如文本、图像和传感器数据）的方法，以增强学习效果和提升决策质量；（4）研究在实时或近实时环境中的协作学习和决策策略，关注动态和不确定环境下的应急响应和关键决策；（5）探究智能体的个性化学习策略，在保持个体优势的同时有效进行集体学习和知识共享。

　　4.多模态融合及生成基础模型。

　　研究多模态数据融合及生成的基础模型，解决数据视角、维度、密度、采集和标注难易程度不同而造成的融合难题；研究模态融合过程中的模态对齐问题，保证模态预测的一致性并减少融合过程中信息损失；研究轻量级的融合模型，提升在模态间非完美对齐状态下融合模型的鲁棒性；研究用易采集、易标注模态数据来引导难采集、难标注模态数据的预训练与微调方法；研究大规模多任务、多模态学习的预训练问题，实现少样本/零样本迁移，发展跨模态多样性数据生成的方法；研究多模态大模型的新型、统一概率建模方法，解决离散、连续混合数据类型的概率建模与生成难题，提高多模态基础模型的生成效率。在多模态模型中实现不少于3个模态的表示学习、对齐及生成能力，模型参数不少于7B，探索在智能座舱、自动驾驶或多模态对话等领域的应用验证。积累用于训练多模态大模型的优质标注数据，并探索数据闭环，采集数量超过标注样本不少于2个量级的非完美标注或无标注数据，实现模型迭代优化。

　　5.模型与数据融合的大模型训练方法。

　　探索系统性的、自适应数据选取方法，以达到数据和模型的有机融合，包括：在模型训练过程中on-the-fly 选取下一步所采用的数据的方法；建立数据和模型有机融合的机器学习框架；发展替代大模型训练普遍采用的先处理数据、再做模型训练的两步走模式的有效方法。

　　6.视频原生的自监督学习方法。

　　针对视频数据既是时间序列又是图像，但又不同于一般的时间序列和图像的特点，充分利用视频数据的属性和特点，发展一套新的自监督学习框架，类比针对时间序列的predict next token 的框架和针对图像数据的填空框架，并在实际视频数据集上加以验证。

　　7.支持下一代人工智能的通用型高质量科学数据库。

　　大规模高质量科学数据是人工智能驱动的科学研究新范式的必要条件。研究科学数据、科技文献等的知识对象标注、抽取、融合中的主动学习机制与自动关联算法；研究面向知识对象的智能编码与机器可识读的多元解析，支持跨领域知识对象的广谱关联，实现与不少于3个国际主流科技资源标识动态互通，支持与外部数据资源智能化融合；研究多模态跨学科知识碎片对齐与知识对象识别方法，以及多学科领域数据自动生产与增强算法，形成符合国际规范或经同行评议的且覆盖不少于8个学科领域的高质量科学数据1PB以上。

　　8.AI for Science 的基础设施建设与示范应用。

　　发展AI for Science的基础设施方法，包括：基础物理模型的人工智能算法；高效率、高精度的实验表征算法；自动化和智能化实验平台建设；科学文献和科学数据的整合与智能应用。发展AI for Science的创新应用，包括但不限于：复杂催化体系（催化剂动态结构变化、反应网络高度复杂等问题）；碳达峰和碳中和中的核心催化反应；工况条件下的电化学表征方法；生物医学中的高效率和高精度的成像技术；有机合成的自动化和智能化解决方案；定向进化蛋白质工程等。重点支持理论和实验相结合并形成闭环的项目。

　　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与支撑。

　　五、2024年度资助计划

　　拟资助培育项目约25项，直接费用资助强度不超过80万元/项，资助期限为3年，培育项目申请书中研究期限应填写“2025年1月1日－2027年12月31日”；拟资助重点支持项目约6项，直接费用资助强度约为300万元/项，资助期限为4年，重点支持项目申请书中研究期限应填写“2025年1月1日－2028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人员不得作为申请人进行申请。

　　（二）限项申请规定。

　　执行《2024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、《2024年度国家自然科学基金项目指南》和《关于2024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2024年4月15日－4月22日16时。

　　（1）申请人应当按照科学基金网络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“培育项目”或“重点支持项目”，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。培育项目和重点支持项目的合作研究单位不得超过2个。

　　（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2024年4月22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4月23日16时前在线提交本单位项目申请清单。

　　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。

　　（四）咨询方式。

　　交叉科学部交叉科学一处

　　联系电话：010-62328382