

2023 年度国家自然科学基金数学天元基金-深圳“数学与智能+”交叉重点专项申请指南

为了推进数学理论、方法与技术医疗健康、电子信息、数字经济等领域中的交叉融合与应用研究，推动数学在战略性新兴产业中的创新发展和应用落地，支撑粤港澳大湾区国际科技创新中心与综合性国家科学中心建设，国家自然科学基金数学天元基金和深圳市科技创新委员会联合设立“数学与智能+”交叉重点专项（以下简称交叉重点专项）。现征集 2023 年度交叉重点专项申请，具体说明和要求如下：

一、科学目标

重点围绕数学与数字经济的“智能+”融合研究，促进应用数学和产业技术创新融通发展。本年度聚焦医疗健康、生命科学等领域，针对医疗健康大模型的构建与应用，开展数学理论、方法与技术的创新攻关研究，以期对重大疾病精准诊疗提供新方法，进一步促进医疗模式创新、优化医疗服务供给，为健康中国战略实施提供理论基础与技术支撑。

二、资助研究方向

本交叉重点专项项目拟资助以下研究方向：

（一）基于随机矩阵极限理论的大模型机理研究。

针对大模型机理研究需求，建立以模型参数规模(P)、训练数据规模(N)、学习质量(如训练模型所达到的损失函数梯度值 ∂l)为三元变量的随机矩阵谱分布极限理论，并用以刻画线性/广义线性大模型泛化性与 $P, N, \partial l$ 之间的变化规律；推导收缩率(Scaling Law)并解释大模型的涌现现象；研究大模型最优（较优）性能的模型规模、数据规模、学习质量之间的最优匹配律，并实验验证；应用线性近似方法或核方法，推广上述结果到非线性情形。

（二）大模型约化的数学理论与方法。

针对大模型的约化开展学习方法论模拟（SLeM）、大模型几何景观、双层优化等创新学习范式与优化理论研究，聚焦于：研究任务的指令化表示与分解体系，构建“任务→方法论学习→方法更新→完成任务”的大模型“单路径”约化方法；研究基于元数据、元知识的大模型约化双层优化理论与高效算法，构建基于 SLeM 理论的大模型约化学习理论与方法；研究文本、图像双模态下的的大模型几何景观性态，提出基于几何景观的大模型约化理论与方法；将所提出的大模型约化理论与方法用于医学多模态大模型，取得显著成效。

（三）支持医学大模型的元数据治理方法与平台。

数据数量和质量是构建大模型的必要条件。为攻关研发医学大模型需要，建立医疗元数据中心平台及医疗病历、影像等多模态医学数据采集、质控、去隐私化、标注的规范和智能技术；联合不少于 50 家国内高水平医院，构建不少于 2 万例头部、胸部、腹部 CT、MR 标准数据集，其中标注数据不少于 1 万例；提出医疗数据隐私保护与分析共享兼备、分层治理与风险优化的数学模型与数学方法；构建基于可信安全计算的多模态医学大模型智能算力平台；验证支持大模型训练和不少于 2 项专科医学智能诊断的下游应用，为医学基础大模型研究提供持续的高质量数据与算力平台支撑。

（注：本项目拟资助经费 300 万元，承担单位需具有对多中心数据的统筹管理资质。）

（四）医学影像判读大模型的研发与关键技术。

基于自主可控基础大模型平台，研发首款医学影像判读大模型。聚焦突破：语言和图像双模态对齐融合、混合推理、图生文、文生图关键技术，医学影像知识增强的语言基础大模型和语言-图像基础大模型，基于增强语言大模型的医学影像自动标注方法，医学影像描述报告自动生成技术，医学影像判读报告自动生成技术等；集成所提出的新技术，研发医学影像自动描述系统和报告自动生成系统，并在 10 家以上三甲医院试点应用。

（注：本项目拟资助经费 300 万元。）

（五）支持医学大模型高效训练的移动计算。

针对医疗数据分散存储所带来的医学大模型训练难题，实现“计算贴近数据”的移动计算范式创新，发展高效实用的分布式学习算法，并支持医学大模型的高效训练和应用。重点研究：基于数据并行、模型分解的无通信、统计高效、无偏的“分解-聚合”分布式训练算法；适应于医学影像多模态处理、异构计算的通信高效类分布式训练算法；带隐私保护医学大数据的高效处理与分析算法；搭建移动计算平台，实现 30P 以上算力的可移动部署，支撑医学大模型的搭建和高效训练及下游应用。

（注：本项目承担单位需配套移动计算平台的硬件经费。）

（六）医学影像大模型的评测与应用模式研究。

构建开源数据、生成数据与专有数据相结合的医学影像大模型评测标准数据集，提出医学影像大模型评测的指标体系与评测方法；研发医学影像“指令-答案”对的智能和渐进标注算法；研究不同威胁级别和不同攻击强度的风险迁移、噪声添加及对抗测试算法。研建端侧和云侧间数据、特征、模型及参数协同链路，实现医学影像端侧个性化推理小模型和云侧基础大模型的协同进化；设计软硬件联合优化

的医学影像大模型蒸馏、微调和增量训练加速算法。面向心血管类疾病，基于 ECG、超声、CT、PET、视频等多源影像数据，研制覆盖报告生成、病灶勾画、干预方案设计、治疗效果预测等核心场景的医学影像诊疗一体化系统，并在 5 家以上大型医疗机构及其医共体开展应用示范。

（七）多模态医学超声基础大模型及其应用。

针对超声影像标准化程度低的突出挑战，研发超声影像判读专用基础大模型。聚焦研究：超声影像的数学化表示与标准化方法；不少于 1 万例的带标注的超声影像标准数据集；超声影像知识增强的语言基础大模型和语言-图像基础大模型；基于指示性强化学习的超声大模型微调高效算法，基于大模型、语义精确可控的阳性病例高维多模态影像合成算法；基于大模型和因果算法的影像自动判读与报告生成数学模型与方法；开发高精度和高泛化性能的超声扫查导航、实时质控和诊断智能系统，并开展产前和心脏领域临床示范应用。

（八）指导蛋白质设计与改造的 AI 模型与系统。

针对蛋白质工程研发全新的兼顾序列和结构信息的预训练模型及下游任务技术需求，建立亿级数据量且面向极端环境标注蛋白序列的核心壁垒蛋白数据库；发展基于粒子系统和消息传递的 Transformer 及 GNN 的数学理论及 AI 模型；用 Transformer 学习蛋白质序列信息，用 GNN 学习蛋白质结构信息，将二者结合研发全新的针对蛋白质工程的预训练框架和下游任务实现 AI 模型与系统。在 10 个以上药用或者工程蛋白上，湿实验验证所开发的系统，实现仅只需少量（<100）突变实验数据就能获得性能优越的满足工程需求的多点位突变的蛋白质产品。

（九）基于人工智能的多肽药物设计大模型。

开发能够针对任意给定靶点蛋白预测活性多肽的 AI 模型。重点研究：如何设计 AI 模型与算法，以快速锁定在指数增长的序列空间中对给定靶点蛋白有药效活性的子空间；如何将 AI 方法与基于第一性原理的计算化学方法结合，提出更加高效的多肽药物设计原理与方法；基于已有蛋白质大语言模型，突破以靶点蛋白作为提示的微调技术，研发专用多肽药物设计 AI 模型，实现在较短时间内将多肽药物的序列空间缩小到湿实验可承受的范围，提升多肽药物研发效率。基于所研发的模型，完成 1-2 个多肽药物的快速设计，验证所研发模型的有效性。

（十）基于三室建模的微重力脑超微结构调控理论与方法。

针对微重力环境下神经兴奋与细胞微环境、血流的偶联作用机制不明的科学问题，应用大数据统计分析和深度学习算法，以不同重力状态下的各结构参数为基础数据集，研究正常重力及微重力下细胞外间隙、血流、脑电信号变化规律，应用微分方程建模与数值模拟方法，研究正向调控技术对超微结构的作用与机制，揭示单室刺激下的三室变化规律，建立微重力条件下神经兴奋功能变化的数学模型；基于

新模型，探索多元素增强的复杂关系表达，构建基于三室模型的神经调控和脑保护新方法，为不同重力条件下的脑保护提供数学技术的新理论、新方法。探索构建基于三室模型及新调控方法的神经系统疾病早期诊断和治疗计划制定的“诊疗一体化”系统，开展基于新理论和新方法的临床试点应用。

（十一）颅内肿瘤多模态数据融合的可通用诊疗模型及系统。

可通用疾病诊疗大模型面临多模态数据融合、资源受限时的基础模型微调和多病种多任务优化等挑战。本项目面向颅内肿瘤，拟开展多癌种、多模态可通用诊疗大模型的构建、微调和评测应用研究，包括：研究医学影像、病理图像、高通量测序和报告文本等多模态数据对齐和融合方法，利用至少 6 万例颅内肿瘤数据构建生物学知识融入的影像-病理-基因多模态基础模型；研究基于少量标注数据、资源受限时的基础模型参数高效微调方法，使之适用于诊断分型、预后评估和新亚型发现等多种下游任务；开展多种常见大模型的微调性能评测和多样化临床任务应用研究，探索最优的诊疗大模型微调策略。

（十二）基于多语言大模型微调的中西医结合风湿病诊疗系统。

针对中西医结合诊疗风湿病机理不清问题，构建以多语言(中、英、古汉语)为载体的医学文本数据集，研究适合中文语言特色的大模型微调新方法和新技术；构建医学影像、病理、临床、生物分子多组学等多模态风湿病数据集，研究跨模态特征关联学习方法将各类型信息融入微调过程促进大模型在垂直领域的“智能涌现”；针对风湿病的 2 到 3 个核心证候，对大模型进行提示调优对齐，减低或避免大模型的“偏见”和“幻觉”影响，形成新一代中西医结合风湿病诊疗系统，并实现示范应用。

（十三）基于交互反馈和领域经验的大模型持续优化与专病辅助诊疗。

面向典型疾病，基于不少于百万名患者的跨模态真实医疗数据，研发支持辅助诊疗的大模型应用系统。系统能在真实诊疗环境下，实现患者精准诊疗方案的计算机智能决策优化；可动态学习由临床专家就个体患者治疗效果与康复进程动态提供的多类型领域经验与多模态反馈，并自适应学习对于不同疾病史患者的个体化最佳诊疗决策。研究与提出在保证患者隐私与安全的前提下，基于真实世界海量异质医疗数据学习，能稳定可靠做出高质量诊疗决策的大模型系统；研究与提出支持上述能力的多模态计算机辅助诊疗大模型持续改进理论及相应的基础数学技术；研究与提出基于交互反馈和多模态领域经验的大模型高效优化理论与增量学习方法。应用上述大模型系统，在不少于五种专病领域，开展真实世界的辅助诊疗应用验证，并在互联网环境下开展覆盖不少于两百万人口的促进普惠医疗和三级诊疗的区域性示范应用。

（十四）基于核医学影像分析/判读的智能诊疗一体化基础模型与系统。

研究针对医学图像学习和推理规则的内蕴关联，建立大模型在医学影像的迁移、可解释性和泛化的模式及机理，构建多模态核医学影像的通用特征表示模型和基于大模型的核医学影像自监督学习范式；提出基于 PET 影像的全身器官高精度分割、配准和病灶快速识别方法；探索 PET 与 CT/MR 相互校正和融合的算法，得到基于影像的定量药物放射性活度，开展关联核素药物摄取和生物效应的图像分析，建立模型和数据双驱动的药代动力学分析方法；探索基于多源医疗数据的肿瘤自动筛查、诊断、疗效评估的算法，基于万例以上多中心核医学影像大数据，构建基于核医学影像的病灶识别、诊断和治疗计划制定的“诊疗一体化”系统，开展基于核医学影像的治疗规划系统（TPS）的试点应用。

三、资助方式与资助计划

以重点项目群的方式资助，项目资助周期不超过五年。项目执行期前两年由数学天元基金与深圳市联合资助（侧重于理论与方法，每项 200 万元，有特别注明的项目，每项 300 万元）；对于明确在深圳市转化落地的项目，后三年拟由深圳市资助（侧重于技术与应用落地，每项不低于 300 万元）；项目资助采取淘汰机制，执行两年后进行中期评估，评估优秀的项目可获连续资助。项目研究团队须由包含数学、医学、信息等不同领域的专家组成，采取双负责人制（其中，排名第一负责人为项目总体负责人）。

2023 年拟资助不超过 14 项。申请书中的研究期限应填写为：**2024 年 1 月 1 日至 2025 年 12 月 31 日**。

四、申请要求及注意事项

（一）申请条件。

本重点专项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的人员不得作为申请人进行申请。

（二）限项申请规定。

1.本重点专项项目不计入高级专业技术职务（职称）人员申请和承担总数 2 项的范围。

2.本重点专项项目申请人和参与者只能申请或参与申请上述十四个研究内容之一的项目。

3.申请人同年只能申请 1 项重点专项项目。

（三）申请注意事项。

1.本重点专项项目试行无纸化申请，项目合作研究单位数量不得超过 2 个。

2.申请人在填报申请书前，应当认真阅读本项目指南和《2023 年度国家自然科学基金项目指南》中申请须知的相关内容，不符合项目指南相关要求的申请项目将不予受理。

3.申请书研究内容应和本指南资助研究内容一致，项目名称要求选择上述十四个研究方向目之一，否则将不予受理。申请书资助项目类别选择“数学天元基金项目”，亚类说明选择“数学天元基金”，附注说明填写“‘数学与智能+’交叉重点专项”。在正文的最前面标明所选研究方向的序号及标题。所有项目申请代码 1 均应选择数学学科申请代码。以上选择不准确或未选择的项目申请将不予受理。

4.数学天元基金项目无间接费用，申请经费为直接费用。申请人应根据《国家自然科学基金资助项目资金管理办法》的有关规定，以及《国家自然科学基金项目资金预算表编制说明》的具体要求，按照“目标相关性、政策相符性、经济合理性”的基本原则，认真编制《国家自然科学基金项目资金预算表》。

5.申请人完成申请书撰写后，在线提交电子申请书及附件材料。申请材料中所需的附件材料（有关证明材料、审批文件和其他特别说明要求提交的纸质材料原件），全部以电子扫描件上传。

五、联系方式

1.填报过程中遇到的技术问题，可联系国家自然科学基金委员会信息中心协助解决，联系电话：010-62317474。

2.其他问题可咨询国家自然科学基金委员会数理科学部数学科学处：

联系人：赵桂萍

电 话：（010）62327191

邮 箱：zhaogp@nsfc.gov.cn