新一代人工智能国家科技重大专项
2025年度第一批项目申报指南

为加快实施新一代人工智能国家科技重大专项（以下简称“重大专项”），新一代人工智能国家科技重大专项专项办公室（以下简称“专项办”）组织编制了2025年度第一批项目申报指南。

重大专项的总体目标是：构建可持续发展的通用人工智能技术体系，创新组织管理机制，持续推动高质量自主可控、安全可信的通用人工智能的基础原始创新和关键技术攻关，为国家安全、人民生命健康和经济可持续发展提供新动力，在未来人工智能领域国际科技竞争中抢占先机。

2025年度第一批项目申报指南以通用人工智能的自主可控、通专融合和安全可信为目标开展基础创新和技术攻关，设置15个研究任务（包含9个青年科学家项目），执行期2-3年，拟安排国拨经费概算9.45亿元。本批研究任务采用“赛马制”组织模式进行部署，通过公开选择先行支持3家单位，在实施一定时间后进行评估择优，确定一定数量的项目继续支持。

重大专项鼓励充分发挥地方和市场作用，强化产学研用紧密结合，调动社会资源投入新一代人工智能研发；同时，鼓励申报团队基于国产基础软硬件开展技术研发和示范验

证。除青年科学家项目不需配套经费外；其他项目均需按指南中相应的配套经费与国拨经费比例进行配套。

测评环节是各研究任务的里程碑检查、赛马遴选和综合绩效评价等过程管理阶段的重要组成部分，具体包括测试方案编制、项目（课题）测评和出具测评报告。基于指南测评思路，主责单位遴选并组织专家编制测试方案，原则上由具有资质的第三方测试机构完成项目测评并出具测评报告。

1.自主可控技术方向

1.1 超大带宽域驱动的大模型架构与系统协同优化

研究内容：研究高带宽、低成本、低功耗的光交换芯片和光模块，基于低成本互联硬件构建高效能超节点。研究高效、通用的数据中心级超节点互联拓扑；研究支持内存语义的统一超节点通信接口；研究超节点和数据中心网络高效配合技术，优化整体通信性能。构建基于光交换的超节点故障隔离、拓扑重配置机制，最小化故障恢复所需的时间。构建面向光交换超节点的训推系统。研究光交换超节点的训练并行策略优化；研究光交换超节点的推理优化，将 PD/AF 分离与超节点相结合；构建支持多款国产芯片的通信库。研究模型-系统-芯片协同优化技术。设计与异步数据引擎深度耦合的高性能模型架构；研发低精度量化友好的模型适配技术，使能 FP8/FP4 超低精度稳定推理；开发面向硬件稀疏计算单元的结构化训练框架。

考核指标：

支持光交换的光模块支持 1ms 以内光路切换，带宽不小于 400Gbps，每 100Gbps 吞吐功耗低于 2 Watts/h，相比 NVL72 系统，单位互联成本降低 25%以上。基于至少一款国产芯片搭建百卡级光交换超节点，峰值 BF16 计算性能不小于 53PFLOPs，每卡互联带宽不低于 1.6Tbps；基于超节点和数据中心网络协同，数据中心网络跨 ToR 交换机流量低于 5%。在 5%服务器故障率情况下健康计算资源的可用率大于 99%；故障造成的计算资源浪费率比 TPU 低一个数量级；支持最优拓扑秒级计算，分钟级拓扑重配置和故障恢复。支持主流模型的高性能训推；典型预训练任务 FLOPs 利用率达到现有非超节点系统的 120%以上；实现基于超节点的 PD/AF 分离机制，推理 TTFT 不大于 5s，TPOT 不大于 75ms。适配光交换超节点特性的通信库发布在国家级开源社区。基于参数量不少于 200B 的 MoE 模型完成超节点训推验证，支持超低精度稳定推理，核心任务精度损失小于 1.5%；通过芯片异步数据引擎和稀疏计算引擎的功能对训推性能深度优化，小批量推理 TPOT 低于 20ms。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

配套经费与国拨经费比例：4:1。

1.2 端侧高性能具身智能 SoC 芯片及工具链研发与规模

应用

研究内容：针对具身智能模型与算法对端侧硬件算力、能效不断提升的技术需求，研究面向具身智能系统任务的高吞吐、超异构计算架构与集成方法，构建基于先进工艺的大算力、高能效具身智能计算架构；研究异构计算架构与具身智能任务的适配技术，提升芯片在端侧推理和增量训练场景的效率和实时性，满足复杂环境下的低延迟响应；研究算法与硬件资源的协同优化技术，集成视觉、语言、运动控制专用加速单元，提升模型推理效率与硬件利用率，实现从算法原型到硬件部署的无缝衔接，满足不同场景的算力需求；研究配套工具链的集成开发技术，包括编译器、运行时和驱动API、性能建模工具、性能优化工具、调试工具、异常检测工具等，适配主流具身智能开发框架与操作系统，降低芯片应用门槛与开发成本，支持具身智能算法模型的快速部署与调试。

考核指标：

基于14nm以下自主国产工艺的芯片指标：CPU算力≥200K DMIPS，稠密AI算力≥400TOPS@Int8，实现INT8/FP16/FP8/FP4多数值格式的兼容支持能力，内存带宽≥400GB/s，系统综合能效≥2 TOPS/W；支持Robobrain+ACT、GR00T-N1等多个主流具身智能模型，并达成模型关键指标，例如运行Robobrain性能≥50/token，ACT推理频

率≥30Hz，端到端延迟≤20ms/步。基于 7nm 以下国际先进工艺的芯片指标：CPU 算力≥400K DMIPS；稠密 AI 算力≥560TOPS@Int8，适配 Transformer、CNN 等主流模型；集成 GPU 核，提供≥200GFLOPS 算力，加速可视化渲染等图形密集型任务；集成 MCU≥3 个，支持亚毫秒级别运动控制周期和精度要求；系统内存带宽≥200GB/s，综合算力能效比≥4.3TOPS/W。两款芯片均需要提供软硬件一体化解决方案，集成从算法开发到系统部署的全链路工具，包括 SDK、开发文档、示例代码及仿真工具等；完成 5 类以上具身智能场景验证，应用于 30 家以上国内主流机器人制造厂商。工具链及算法核心模块在国家级人工智能开源社区中开源发布。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

配套经费与国拨经费比例：5:1。

1.3 面向自主超节点的大模型推理软硬协同优化研究（青年科学家项目）

研究内容：针对大模型的模型规模、专家数量和上下文长度持续增长的趋势，面向自主可控超节点集群研究混合并行机制、微服务调度架构、长上下文稀疏推理、异构算力推理等技术，在保障模型精度下提升单位算力或成本下的系统

吞吐量。

考核指标：

给定 1-2 个主流的千亿规模以上开源模型，自主超节点的平均单位算力吞吐率（Throughput per TPLOPS）不低于 0.8，基于自主异构算力的推理系统在单机上实现系统吞吐量不低于 50 tokens/s。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

1.4 物理知识嵌入的注入式复杂场景生成与仿真关键技术（青年科学家项目）

研究内容：针对现有复杂场景仿真中物理真实性不足和生成效率低等问题，研究物理知识嵌入的注入式复杂场景生成与仿真关键技术。构建低重力、弱光照极端环境下，物理规律驱动的高保真虚拟场景生成与仿真系统，支撑极端工况下探测车、协作机器人等智能体的感知、决策与控制算法训练、验证及部署前评估的大规模、多样化复杂场景库，显著提升其在极端场景（低重力，弱光照）中的任务执行能力和鲁棒性。

考核指标：

构建物理嵌入的生成式仿真算法及高保真实虚融合场景库（规模>1 万条）；实现在主流开源技术平台（如桃源 InternUtopia）上的系统集成，支持低重力、弱光照极端环境

下的 10 种以上典型工况仿真；在闭环仿真中，智能体感知准确率不低于 85%，核心任务执行成功率达到 90%，Sim2Real 关键技能迁移性能损失不超过 15%。算法核心功能模块在国家级人工智能开源社区中开源发布。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

1.5 面向具身操作的力触仿真（青年科学家项目）

研究内容：针对目前纯视觉感知策略无法适应复杂操作任务的问题，研究高效的力触仿真方法，构建支持力触感知的仿真训练环境，在基本不影响训练速度的情况下，显著提升复杂操作任务的成功率。

考核指标：

在铰链型刚体和布料柔性体等大规模操作任务的仿真过程标记点位移场与真实之间的最大相对误差不超过 30%，支撑模型训练 sim2real 的闭环验证，操作铰链型刚体与柔顺夹取易碎品等典型应用的任务成功率不低于 90%。算法核心功能模块在国家级人工智能开源社区中开源发布。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

2.通专融合技术方向

2.1 知识增强的科学具身智能体平台

研究内容：构建融合通用大模型与领域知识的 AI4S 科

学具身智能体平台，形成兼具通用性与专业性的科学认知新架构；构建基于 AI4S 科学知识图谱的科学智能体，并融合通用大模型，研究两者间的双向增强与动态更新机制。研发具备慢思考与可控推理能力的科学智能体，实现复杂科学任务的自主理解与实验方案设计；研制嵌入科学机理的数字孪生仿真训练环境，高仿真模拟实验全流程，用于加速科学智能体对高级技能的习得与泛化，并以工具链形式集成到科学发现平台；研发软硬一体的科学装置智能体平台套件，通过多智能体协同，集成并智能操控大规模科学装置。构建实验设计—执行—反馈的闭环路径，支持复杂科研任务的全流程自主执行；聚焦生物化学、生命免疫、材料科学等典型学科，部署并验证自主化实验系统。建立标准化评估基准，检验平台在真实多维任务中的有效性，推动科研范式向智能化、自主化转型。

考核指标：

构建高质量 AI4S 知识图谱，覆盖生物化学、生命免疫、材料科学等不少于 3 个科学方向，规模与处理能力达现有水平 10 倍以上，知识准确率平均不低于 90%；科学智能体具备博士级的知识理解能力，知识自动更新准确率不低于 95%，在典型科学任务场景中，平均幻觉率控制在 5%以下，推理链条可追溯率不低于 85%。开发科学具身智能体虚拟训练系统，实现至少 50 种科学实验流程的高仿真模拟，确保基于

仿真数据训练的智能体策略在零样本迁移至实体机器人时，首次 Sim-to-Real 成功率超过 60%。构建软硬一体的科学具身智能实验系统，具备多机协同执行实验任务能力。支持超过 7 天的全流程自主科学实验,实验规划时间缩短超过 60%，推动科研综合效能提升至少 50%。在生物化学、生命免疫、材料科学等不少于 3 个学科方向的 10 个真实科研任务场景中完成系统部署与验证，在至少 1 个学科方向上形成新理论或新发现。发布科学具身智能体开源平台 1 套，形成关键技术规范与行业示范。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

配套经费与国拨经费比例：3:1。

2.2 具有环境任务泛化能力的通才智能体

研究内容：针对通才智能体在跨硬件适配、任务泛化和极端环境适应中面临的高依赖性、低稳健性和进化不足等核心挑战，研究面向智能体工厂的模块化模型构建方法，实现模型的可重用与可扩展，形成可组合与进化的原子技能库；构建长期自主探索与自适应学习的机制，确保智能体在多种常见或特殊环境中，以及环境剧变与任务突发条件下能够快速恢复并持续优化，实现数小时级的稳定运行；面向智能制造操控、极端环境探索、复杂博弈对抗等典型场景开展应用集成与验证。构建并交付针对各类常见与特殊环境可自主适

应、自主泛化、自主生长通才智能体计算模型与组件栈，打造一套具备方法体系与模块化组件的通才智能体框架，在主流开源技术体系或平台与多型本体设备上集成应用，支撑整体具身智能的构建与规模化推广。

考核指标：

构建不少于5个高复用性核心模型（具身感知/世界模型/决策控制等），形成可组合原子技能库规模≥60项；新环境和新本体的迁移上线时间在常规场景不超过10分钟、在极端场景不超过30分钟，在≥7类环境和≥4类本体平台上稳定执行≥30类任务，对未见任务的首次执行成功率在常规环境中达到≥80%、在极端环境中达到≥70%；在高端应用场景实现突破性验证，智能制造误操作率降至0.1‰，极端环境勘探效率提升25%，复杂博弈决策速度达毫秒级，综合部署成本降低30%；形成的模型、算法等软件成果，将部分可独立运行的核心功能模块开源，优先发布于国内主流人工智能开源社区。通过典型测试场景对标国际领先的具身智能模型，在量化评测上对标领先开源模型，在真实作业能力边界评测上对标各种开闭源模型。

组织方式：公开竞争。可公开选择3家赛马后滚动支持。

配套经费与国拨经费比例：3:1。

2.3 数理约束下的科学序列数据生成技术（青年科学家

项目）

研究内容：针对以科学基础模型训练的 AI-Ready 数据需要，研究数理约束的科学序列数据智能生成框架；研发基于高阶稀疏关联的预测生成方法，构建科学数据的广义标签增量迭代机制，实现科学数据的缺失补全（通道级）、采样对齐（跨频率）和持续生成（十万级）；构建可数值推理的序列数据 Token 化方法，将地球科学、物质科学、生命科学等不少于 5 类科学专用领域的数据扩充为可训练科学语料。

考核指标：

研发一套科学序列数据专用生成工具，实现≥10 万时间节点的科学序列数据稳定持续生成，高保真数据生成的准确率不低于 80%，覆盖天文学、地震学、电磁学、电化学、临床医学等 5 个专用领域；构建数值可推理的 Token 化方法，序列关联保序稀疏度超过 60%，在 5 个专用领域训练模型性能提升超过 20%，支持通专模型的深度集成。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

2.4 基于因果增强的深度推理语言大模型技术（青年科学家项目）

研究内容：针对现有语言大模型深度推理过程中的错误和幻觉严重的问题，研究因果增强的深度推理训练算法，构建可支持复杂因果推理场景的高质量数据集和评测集，在基

本不影响模型通用能力的情况下，显著降低推理错误率与幻觉率。

考核指标：

采用此方案对训练阶段进行优化后的语言大模型，在通用能力基本不降低的情况下，对于教育、司法、医疗等错误和幻觉敏感类领域，推理错误率相对下降不低于30%，幻觉相对下降不低于50%。

组织方式：公开竞争。可公开选择3家赛马后滚动支持。

2.5 面向大模型长程交互的持续自主进化技术（青年科学家项目）

研究内容：针对当前人工智能大模型在长程多轮对话中存在的一致性差、工具使用错误率高及复杂指令遵循能力不足等问题，研究自主进化策略学习与层级任务分解方法，构建面向大模型长程交互的持续自主进化框架，从而支持跨会话、跨任务、跨时间尺度的长程交互和持续自主进化学习。

考核指标：

采用此方案进行持续自主进化训练的大模型，在长程多轮任务完成率较基线系统提升不低于5%，工具使用正确率提升不低于10%，在10轮以上的长对话中前后不一致性错误减少不低于15%。

组织方式：公开竞争。可公开选择3家赛马后滚动支持。

3.安全可信技术方向

3.1 基于数据价值的内容安全治理

研究内容：构建数据价值驱动的治理体系。针对大语言模型在训练与生成过程中的语料来源不明、语义风险难追踪等问题，提出融合语义稀缺度、伦理敏感度、文化忠实度和任务关联度等特征的因果追踪方法，识别高价值语料。研发异常数据识别与追踪机制。面向训练语料中潜在的数据投毒与污染风险，研发可解释的多层次识别与追踪机制，实现对异常干预与恶意注入的快速定位与修正，保障训练数据的纯净性与可信性。构建分级分类的风险治理方法。结合社会治理需求与行业应用场景，研究面向不同语义风险的内容分级分类方法与知识编辑方案，确保模型输出兼顾文化多样性与合规性。开发全链路合规追踪工具。建立覆盖数据输入、训练过程与内容输出的合规追踪体系，支持数据、推理路径及神经元层面的风险可视化审计，提升模型部署的可监管性与安全可控性。

考核指标：

在数据安全规范方面，制定符合社会主义核心价值观的数据治理标准体系，建设典型样例库不少于5万条、语料库规模不低于10亿条，形成高质量语料基础。在功能性能方面，系统具备多语言语料的语义权重建模与风险分层能力，

覆盖不少于三类社会治理典型场景，数据权重判别准确率≥90%，污染样本识别率≥95%，合规预警覆盖率≥95%，误报率≤5%，高风险语义内容减少≥30%。在成果应用方面，完成安全治理原型平台的部署与实际验证，落地于至少 1 个自主基础大模型和 10 个行业大模型，整体合规性水平提升≥25%，形成可推广的工具链和规范框架。在开源贡献方面，将多特征多层次的因果追踪、多层次风险识别模块等在国家级人工智能开源社区中开源发布。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

配套经费与国拨经费比例：3:1。

3.2 多模态大模型的内生安全

研究内容：针对多模态大模型在跨模态对齐与语义融合过程中易出现的模态错配、幻觉失真、价值冲突等问题，构建逻辑、事实、语义、意图约束的多维度可信推理框架，研究跨模态可验证推理链，提升推理过程的透明性与可解释性。研究多模态生成幻觉检测与评估技术，设计跨模态生成不确定性量化方法，对多模态大模型进行风险评估与校准，降低模型幻觉问题。研究多模态大模型价值对齐与纠正方法，实现价值嵌入与行为约束一体化，实现模型输出结果的多样、公平与合规。针对多模态大模型内生有害知识与后门问题，研究跨模态遗忘学习与动态参数修复算法，使模型具备主动

遗忘、快速净化与自我修复能力。构建覆盖多模态大模型全生命周期的内生安全框架。

考核指标：

实现安全基准测试 SIUO 通过率≥90%，跨模态逻辑一致性检测准确率≥90%，输出幻觉率≤5%；提升跨模态推理校准性能 20%以上，并在实际场景中实现≥85%的高风险内容提前预警准确率；跨模态价值一致性保持率≥95%，违规内容发生率降低≥60%，同时正常内容生成能力≥95%；开发自适应遗忘与内生演化机制，实现有害知识清除率≥85%，后门投毒样本检测成功率≥90%，防御能力提升≥30%，并保持演化场景下≥80%的性能稳定性。在示范应用方面，在多个业务部门的示范平台进行实际部署与验证，技术就绪度 TRL 达到 7 级以上。在开源贡献方面，将跨模态可验证推理链、价值对齐与纠正方法模块等在国家级人工智能开源社区中开源发布。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

配套经费与国拨经费比例：3:1。

3.3 个性化智能体的可信推理与安全执行技术（青年科学家项目）

研究内容：针对个性化交互场景中普遍存在的风险感知不足、行为易受操纵以及防御与个性化体验难以平衡等问题，

研究并开发即插即用的可信推理与安全执行框架，全面增强自主安全感知、动态防御和可验证推理链能力，使智能体能够自动识别潜在风险、及时调整策略、抵御外部操控攻击，并确保决策过程透明与合规，整体性能接近或达到国际最优水平。

考核指标：

采用此方案对大模型智能体，实现个性化交互场景下通用能力基本不降低的情况下，安全可信能力提升不低于30%，安全防御机制与任务性能的协同优化能力达到国际先进水平，支撑个性化推荐交互、用户界面交互、对话交互等三类代表性场景的可信应用部署。

组织方式：公开竞争。可公开选择3家赛马后滚动支持。

3.4 具身感知的安全可信技术（青年科学家项目）

研究内容：针对开放复杂环境中具身感知模型本体概念弱、与物理操作关联度低、虚实世界差异大、难以在线感知进化等问题，研究具身感知模型本体增强方法、感知与操作容错增强方法，增强虚实泛化迁移能力，赋予具身感知模型在线安全进化能力，实现从原始传感数据到环境理解的全链路安全可控与透明合规，同时保持通用性能接近或达到国际最优水平。

考核指标：

采用此方案对典型具身大模型进行训练与优化，在核工业、高危化学实验操作环境等领域典型应用场景的开放复杂环境中进行评测，实现通用感知性能基本保持（精度变化不超过 1%）的前提下，具身大模型的风险行为拒答能力提升不低于 30%，复杂任务成功率提升不低于 40%。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。

3.5 面向具身多模态大模型的安全可信后训练技术（青年科学家项目）

研究内容：研究面向具身多模态大模型的安全可信后训练技术。基于具身智能可扩展大规模强化学习框架，提升模型对不安全指令的拒答能力，通过强化学习后训练实现具身智能体自主安全进化；提升具身大模型在开放指令空间下的行为合规性与复杂操作可控性；构建生成理解统一具身多模态大模型，通过对未来预测规避风险提升模型安全。

考核指标：

设计一套具身多模态大模型的安全可信后训练框架，实现具身大模型的高效安全复杂推理，训练效率提升不低于 50%，模型风险率降低 30%，实现具身智能体自主进化，使模型对不安全指令的拒答能力提升 30%以上，模型成功率提升 40%以上。提交一套来自真实世界的、具有 3 种以上模态交互场景的具身大模型通用安全评测集，覆盖 3 类以上机器

人本体（轮式、足式、灵巧手），完成 50 项以上高复杂度任务验证。

组织方式：公开竞争。可公开选择 3 家赛马后滚动支持。