

大数据与决策国家级重点实验室

# 2026 年度开放基金指南

2026 年 9 月

# 目 录

|                                             |    |
|---------------------------------------------|----|
| 课题 1: 可自主演化的智能标注关键技术研究 .....                | 1  |
| 课题 2: 多模态图文数据标注与数据集构建方法研究 .....             | 2  |
| 课题 3: 基于智能体的跨域数据融合与可信评估方法研究 .....           | 3  |
| 课题 4: 面向复杂数据治理任务情景的人因工程稳健性研究 .....          | 5  |
| 课题 5: 基于大语言模型的数据治理规则自动生成研究 .....            | 6  |
| 课题 6: 面向图像目标检测的高质量数据集评测方法研究 .....           | 8  |
| 课题 7: 多模态数据敏捷主动服务技术研究 .....                 | 9  |
| 课题 8: 面向智能数据分析的 Agent 自演化关键技术研究 .....       | 11 |
| 课题 9: 基于多智能体协同的数智共融方案生成与可信协同决策研究 .....      | 12 |
| 课题 10: 指挥决策多模态本体构建与编排技术研究 .....             | 13 |
| 课题 11: 基于 Hybrid 智能体架构的多模态事件本体知识库构建技术 ..... | 16 |
| 课题 12: 潜在因果识别驱动的持续演化事件图构建方法研究 .....         | 17 |

## 课题 1：可自主演化的智能标注关键技术研究

### 1. 研究目标

针对智能标注模型难以在人机交互中持续进化、无法高效纠正错误积累的挑战，通过设计人机协同标注交互模式，突破基于持续学习与主动遗忘的标注模型动态优化等技术，实现智能标注模型“越标越好、有错即改”能力，为构建自进化、高可信的数据生产体系提供新途径与技术支持。

### 2. 研究内容

（1）针对现有人机标注流程割裂、协同效率低的问题，设计高效协同、可动态调整的人机协同标注流程，解决如何根据专家标注实时优化机器标注流程的难题。

（2）针对标注模型性能停滞问题，研究增量式学习方法，实现在标注过程中模型能够吸收专家标注知识，提升自动标注精度，而不产生灾难性遗忘问题。

（3）针对机器标注产生的系统性、重复性错误，研究基于错误反馈驱动的选择性遗忘学习方法，实现快速、精准地遗忘错误信息关联，而不损害已有正确知识。

（4）结合图像目标识别、文本实体抽取等典型应用场景，开展人机协同智能标注案例设计与分析。

### 3. 主要技术指标

（1）支持图像目标识别、文本实体抽取等任务的数据自

动标注;

(2) 典型任务下, 模型自动标注准确率  $\geq 85\%$ ;

(3) 模型经历 10 轮迭代, 对旧数据的性能遗忘率  $\leq 10\%$ , 对新数据的平均标注精度提升  $\geq 20\%$ ;

(4) 在接收到错误标注反馈后, 模型在 5 轮迭代后对指定错误模式的召回率下降至  $\leq 5\%$ 。

#### 4. 成果形式

研究报告、算法模型。

#### 5. 经费周期

2 年, 25 万。

### 课题 2: 多模态图文数据标注与数据集构建方法研究

#### 1. 研究目标

针对图像语义层次复杂、场景多样、信息不统一等特点, 研究图像与文本之间的语义关系, 突破图像多模态语义描述不一致、数据标注成本较高等关键技术问题, 设计多粒度图像—文本标注规范, 构建规模适中、语义清晰、结构规范的多模态数据集, 形成可复用的数据集构建流程与标注规范。

#### 2. 研究内容

(1) 针对图像在典型应用场景下平台类型多样、呈现形式差异明显的问题, 研究基于平台类型和场景特征的数据整理与分类方法, 形成可复用的数据组织与整理流程。

(2) 针对图像语义表达不统一的问题, 研究多粒度图像

一文本标注规范，设计全局描述、语义描述及基础结构化语义字段。

(3) 针对数据标注成本高、质量难以保证的问题，研究模型辅助生成与人工校正相结合的标注流程，并开展一致性校验。

(4) 结合标注规范与标注流程，构建特定场景多模态图像—文本数据集，研究数据集质量评估方法，并对数据质量和统计特性进行分析验证。

### 3. 主要技术指标

(1) 构建特定领域图像—文本多模态数据集规模不少于 5000 条；

(2) 文本标注至少包含 3 种不同语义粒度描述信息；

(3) 文本描述与图像内容语义一致率不低于 90%，并通过人工抽样核验；

(4) 构建的数据集应采用结构化、通用的数据组织格式，如：JSON、CSV 等。

### 4. 成果形式

研究报告、多模态图文数据集。

### 5. 经费周期

2 年，20 万。

## 课题 3：基于智能体的跨域数据融合与可信评估方法研究

### 1. 研究目标

聚焦应急事件处置全域智能演进需求，研究基于智能体的跨域数据融合与可信评估方法，实现跨域数据的可信互联与一致性协同。

## 2. 研究内容

（1）面向跨域场景下的异构数据关联对齐方法。针对跨域数据模态异构、语义差异大、关联关系模糊的问题，研究多模态异构数据的表征与自动化关联对齐方法，实现跨域数据一致性协调融合。

（2）基于智能合约的多域数据可信监管机制。针对跨域数据身份互信难、流程不可审计、数据主权易丧失的问题，利用去中心化网络的智能合约技术构建多域数据可信监管机制，实现数据流通流程可信、可靠与可追溯审计。

（3）基于智能体的数据动态态势可信评估与优化方法。针对传统评估模型僵化、数据可信度难量化、评估结果缺乏反馈优化的问题，结合智能体技术设计动态跨域数据可信评估算法，提升评估的自适应性能力和精准度。

## 3. 主要技术指标

（1）异构数据关联对齐方法涵盖的跨域数据来源领域不少于3种，可支持分析的结构化与非结构化数据类型不少于4类；

（2）在跨域数据协作场景，数据流通全过程审计留痕，可在分钟级完成异常事件的跨主体审计追溯与责任定位；

(3) 动态跨域数据可信评估算法模型支持分钟级( $\leq 5$  分钟)的完整更新与推理周期。

#### 4. 成果形式

研究报告、算法模型。

#### 5. 经费周期

2 年, 20 万。

### 课题 4: 面向复杂数据治理任务情景的人因工程稳健性研究

#### 1. 研究目标

面向复杂数据治理情景中任务驱动的人机交互界面结构稳定性问题, 围绕界面可操作单元在复杂任务条件下形成的操作网络结构, 研究其关键结构特性、抗扰动机理及结构重构规律, 提出界面结构稳定性评估与优化机制, 形成面向复杂任务的界面可操作单元结构建模、分析与评估方法体系。

#### 2. 研究内容

(1) 面向复杂任务驱动下界面可操作单元之间结构关系不清、关键单元难以识别的难题, 研究基于层级任务分析与复杂网络相结合的界面结构建模方法, 将任务需求、操作行为与界面单元统一映射为操作网络, 通过网络拓扑特性分析, 识别界面中的关键操作单元与核心交互路径。

(2) 针对复杂任务条件下界面结构稳健性缺乏定量评估手段的问题, 应用复杂信息系统网络抗毁熵测度方法分析网络鲁棒性, 定量评估界面结构在局部失效与整体扰动条件

下的稳定性。

(3) 提炼面向复杂数据治理任务系统界面的结构化设计与优化思路，形成可推广的方法框架，解决界面设计与调整依赖经验、缺乏可迁移优化方法的难题。

### 3. 主要技术指标

(1) 界面结构稳健性指标：典型复杂任务情景下，相较于原始界面结构，优化后人机交互界面的结构稳健性指标提升 20%以上；

(2) 关键操作单元抗扰动能力：在模拟关键操作单元失效或操作负载突增条件下，优化后界面结构的任务可达性保持率不低于 85%，明显优于对照界面结构；

(3) 界面操作准确性：在复杂任务仿真环境中，数据中台人机界面操作准确率达到 80%以上，关键操作步骤错误率显著下降。

### 4. 成果形式

研究报告、算法模型。

### 5. 经费周期

2 年，20 万。

## 课题 5：基于大语言模型的数据治理规则自动生成研究

### 1. 研究目标

传统的数据治理依赖人工经验预先定义规则，难以适应动态变化的数据治理场景。通过大模型的语义理解与推理能

力，实现治理规则和流程的“自动生成”，从而降低数据治理的使用门槛与人力成本，提升数据治理的整体效率。

## 2. 研究内容

(1) 数据治理规则的存储模型体系构建。基于向量数据库设计兼具灵活性与可扩展性的存储方案，建立相应的分类体系，形成可迭代积累的规则知识库，实现对已有人工构建数据治理规则的高效管理。

(2) 基于用户意图理解的治理目标分类。将用户的数据治理需求文本，转换为机器可识别可理解的信息，并对用户提出的治理问题进行自动分类，为后续智能化数据治理提供支撑。

(3) 基于用户关键词抽取的元数据匹配。研究基于用户关键词抽取的元数据匹配技术，实现治理规则与元数据的智能匹配，解决用户输入内容与数据源字段难以匹配对齐的问题，为基于大语言模型的数据治理规则脚本的自动生成提供关键参数。

## 3. 主要技术指标

(1) 根据用户语义识别生成相应治理规则的类别>5个；

(2) 基于用户输入进行数据治理任务分类的准确率>80%；

(3) 自动生成的数据治理规则准确率>80%。

## 4. 成果形式

研究报告、算法模型。

## 5. 经费周期

2 年, 20 万。

# 课题 6: 面向图像目标检测的高质量数据集评测方法研究

## 1. 研究目标

聚焦图像数据集质量评测需求, 建立多维分层质量评测理论模型与指标体系, 突破时序特征量化关键技术, 研究适配动态目标及静态目标的评测技术, 提出典型目标的场景化评测方法, 研究标准化的评测实施流程, 支撑数据集评测自动化、规范化。

## 2. 研究内容

(1) 大场景多目标图像数据集多维分层质量评测体系研究。系统设计多维度、差异化的质量测评指标体系, 重点针对无人机、车辆、舰船等典型目标在不同距离与尺寸下的成像特性及在目标检测识别领域场景的应用需求, 制定核心指标与判定标准, 明确各分层组合的指标阈值与计算方法。

(2) 时序数据集运动特征与静态特征质量评测技术研究。面向时序视频数据集, 构建动态与静态特征协同评测的技术体系。针对动态目标, 研究运动模糊度、帧间同步性等量化评测技术; 针对静态目标, 研究帧间位置稳定性、分割轮廓一致性等测评方法。通过区分动态与静态目标的特性差异, 弥补现有测评对时序关联特性与静态稳定性关注不足,

确保测评结果覆盖数据的全维度质量特征。

(3) 标准化实施流程设计。制定测评标准化实施流程，设计统一报告模板，解决评测流程不规范、结果不可复现的问题。

### 3. 主要技术指标

(1) 构建多维指标框架，覆盖采集、标注、内容、适配性等类别，核心量化指标总数不少于 15 项；

(2) 针对小不同分层下的样本失衡、标注偏差等典型缺陷，指标灵敏度  $\geq 85\%$ ；

(3) 同一数据集在不同运行环境下，重复计算同一指标的结果偏差  $\leq 2\%$ ；

(4) 时序同步性评测能力，时间戳异常检测准确率  $\geq 90\%$ ；

(5) 标准化实施流程方案需完整覆盖测评全生命周期，包含不少于 6 个核心环节的操作标准、输入输出物定义。

### 4. 成果形式

研究报告、算法模型、测评标准化实施流程方案。

### 5. 经费周期

2 年，25 万。

## 课题 7：多模态数据敏捷主动服务技术研究

### 1. 研究目标

针对多源异构数据融合僵化、跨模态语义对齐失真、交

互信号可信度低的挑战，开展多模态数据敏捷主动服务技术研究，突破差异感知分层融合、双空间表征学习、交互图净化与多图协同学习等关键技术，构建自适应多模态主动服务模型，实现用户需求精准刻画与信息敏捷精准推送。

## 2. 研究内容

(1) 针对多源数据模态差异大、融合策略单一的问题，提出基于差异感知的分层融合机制，量化模态间分布差异，动态结合早期与后期融合，实现多源信息有效集成互补。

(2) 针对跨模态特征对齐易丢失专有语义信息的问题，构建全局对齐、局部保留的双空间表征学习框架，提取共性特征的同时保留各模态独有细节，提升用户需求刻画精度。

(3) 针对复杂场景下交互信号含噪高、易受虚假信息干扰的问题，设计交互图净化与多图协同学习方法，抑制误导性交互，构建可信的用户信息交互网络，保障信息精准推送与服务。

## 3. 主要技术指标

(1) 能够处理的多模态信息种类不低于 2 种；

(2) 建立至少 3 个多模态数据敏捷主动服务算法模型；

(3) 针对多模态数据主动服务准确率，与传统方法相比，效果提升不低于 15%。

## 4. 成果形式

研究报告、算法模型、专利。

## 5. 经费周期

2 年，20 万。

## 课题 8：面向智能数据分析的 Agent 自演化关键技术研究

### 1. 研究目标

针对当前自动化数据分析与建模系统中普遍存在的短视指标陷阱、统计严谨性缺失以及面对非标脏数据时流水线易崩溃的痛点，以“安全边界内的自主涌现”为核心建设思路，构建一种自下而上代码级涌现与自上而下统计语义规训相融合的混合自演化智能体框架，确保数据分析智能体在强对抗、高动态环境下的可靠性与可控性，为大数据全自动分析提供底层算法支撑。

### 2. 研究内容

（1）针对领域数据分析“零先验”困境，研究基于零先验的自下而上数据处理技能涌现与动态建库机制。

（2）针对领域无法接受“黑盒”决策，研究基于大语言模型语义审计的全局步骤级归因机制。

（3）针对在长期在线学习中，智能体可能为了短期奖励最大化而采取投机取巧的策略，导致长期效能下降，研究双通道复合奖励与强制经验混合探索机制。

（4）结合以上研究内容，设计一种自下而上代码级涌现与自上而下统计语义规训相融合的混合自演化智能体框架方案。

### 3. 主要技术指标

(1) 冷启动收敛时间：在零先验条件下，针对新型非标准数据格式，Agent 自主涌现出有效特征提取与清洗技能的时间  $\leq 30$  分钟；

(2) 步骤级故障归因准确率：当分析结果出现偏差时，能精准定位到具体出错的数据处理步骤的概率  $\geq 92\%$ 。

### 4. 成果形式

研究报告、算法模型。

### 5. 经费周期

2 年，20 万。

## 课题 9：基于多智能体协同的数智共融方案生成与可信协同决策

### 1. 研究目标

围绕“数据-智能体”分离导致方案生成动态性不足、决策可信度低的核心难题，研究多智能体通过结构化交互对多源异构数据进行动态对齐、关联与价值挖掘的内在机理，以及多模型协同迭代中缓解幻觉、形成共识的博弈演进规律，突破“数据-知识-决策”的实时融合技术、面向任务的多模型动态调度与协同、多模型协同博弈推理等关键技术，实现从静态数据支撑到动态数智共融的转变，提升专业大模型从信息处理、分析到决策的协同效率。

### 2. 研究内容

(1)多模型动态组织与调度机制。针对任务复杂多变、专业模型异构分散导致单一模型决策困难的问题，构建动态模型调度中枢，实现任务的高效分解与多模型动态调度。

(2)多模型结构化协同决策与共识达成推理技术。针对单一模型存在固有局限与“幻觉”、群体交互易陷入低效争论或共识偏移的难题，研究多智能体协同决策与推理技术，实现多模型有效共识推理决策。

(3)数智共融驱动的多模型协同决策技术。针对数据、多智能体难以高效融合导致数据赋能智能困难的问题，开展数据与多模型融合决策研究，依托专家数据提高多模型协同决策能力。

### 3. 主要技术指标

- (1) 任务分解准确率  $\geq 95\%$ ;
- (2) 协同决策大模型数量  $\geq 3$  个;
- (3) 协同决策相比单一大模型的错误率下降  $\geq 30\%$ ;
- (4) 结合专家数据后的协同决策错误率下降  $\geq 50\%$ 。

### 4. 成果形式

研究报告、算法模型、专利。

### 5. 经费周期

2 年，20 万。

## 课题 10：指挥决策多模态本体构建与编排技术研究

### 1. 研究目标

复杂场景下数据呈现多源、多模态、多尺度、高动态特征，传统“数据一查询一展示”模式难以支撑智能化指挥决策需求。本体建模可将异构数据抽象为统一“对象一属性一关系一动作”语义，为跨域数据融合与知识沉淀提供路径，但仍面临动态编排不足、语义与执行脱节、复杂场景下可靠性不足等问题。项目拟构建“语义感知、智能规划、自主编排”的本体分析体系，推动指挥决策由“数据驱动”迈向“语义驱动”。

## 2. 研究内容

（1）针对复杂指挥决策场景下异构数据统一语义抽象难题，研究面向多模态本体建模与操作代数构建方法，建立“多模态数据一本体语义一计算原语”三层映射关系，形成可编译为原语 DAG 的本体操作体系。

（2）针对复杂本体分析任务动态规划难题，研究智能体驱动的本体任务自主规划与分解方法，融合大模型推理能力与领域知识，支持从自然语言指挥意图或结构化本体调用到计算原语 DAG 的智能分解，实现可解释、可校验、可修正的规划输出。

（3）针对本体操作涉及多类异构计算组件时的执行协同难题，研究本体语义感知的跨组件任务编排与执行优化技术，结合实时算网状态、数据分布和组件负载信息，优化执行位置、执行顺序与资源配额，并支持跨模态算子融合执行。

(4) 结合指挥决策场景对可靠性、抗毁性和可干预性的要求，研究多智能体协同决策与闭环执行机制，构建规划智能体、编排智能体、执行智能体、反思智能体协同架构，支持异常感知、动态重规划、降级执行和人在回路干预。

### 3. 主要技术指标

(1) 构建指挥决策领域本体，支持核心本体关系类型  $\geq 30$  类、本体动作语义  $\geq 20$  类；本体操作代数支持计算原语  $\geq 15$  种，覆盖图、时空、向量、关系、流式等不低于 5 类计算模态；本体操作代价模型预估误差  $\leq 15\%$ ；

(2) 复杂本体分析任务自动分解成功率  $\geq 90\%$ ，规划结果可执行性验证通过率  $\geq 95\%$ ；典型任务规划响应时延  $\leq 5$  秒，支持规划意图类型  $\geq 10$  类；

(3) 相比静态调度基线，典型本体分析任务端到端执行时延降低  $\geq 20\%$ ；跨组件算子融合执行支持复合操作模式  $\geq 8$  类；多模态计算组件统一编排支持  $\geq 5$  类主流组件；

(4) 构建规划—编排—执行—反思四类核心智能体角色，定义协同交互协议消息类型  $\geq 10$  种；支持异常类型识别  $\geq 8$  类，降级执行策略  $\geq 3$  级，人在回路干预覆盖规划、编排、执行三个层级。

### 4. 成果形式

研究报告、算法模型、专利。

### 5. 经费周期

2 年，25 万。

## 课题 11：基于 Hybrid 智能体架构的多模态事件本体知识库构建技术

### 1. 研究目标

面向应急事件处置中的事件本体构建问题，开展基于 Hybrid 智能体的多模态事件本体特征发现与动态本体知识库构建研究，研究多智能体协同与动态知识库实时演进机制，突破基于 OODA 环的事件本体模型构建、基于多模态感知智能体的事件本体特征发现和基于认知智能体的动态事件本体知识库构建等关键技术，构建混合智能体多模态事件本体模型，完成事件本体构建的验证测试。

### 2. 研究内容

（1）针对事件本体模型体系化与粒度层次不清晰的问题，开展基于 OODA 环的事件本体模型构建技术研究。

（2）针对多模态数据中本体信息获取自动化智能化能力不足的问题，开展多模态感知智能体的事件本体特征发现技术研究。

（3）针对事件本体自动演化迭代的技术瓶颈，开展基于认知智能体的动态事件本体知识库构建技术研究。

### 3. 主要技术指标

（1）事件本体模型不少于 8 个，同时需包含不同粒度层级，属性维度不少于 10 项；

(2) 智能体可识别文本、图像、表格等模态的数据，关键本体目标识别率不低于 90%；

(3) 智能体可结合事件本体模型进行文本、图像模态的本体统一构建与关联关系发现；

(4) 智能体任务完成率不低于 90%，调用成功率不低于 85%。

#### 4. 成果形式

研究报告、算法模型、事件本体模型。

#### 5. 经费周期

2 年，20 万。

### 课题 12: 潜在因果识别驱动的持续演化事件图构建方法研究

#### 1. 研究目标

面向应急事件动态演化与因果关联复杂性所带来的态势分析评估挑战，开展潜在因果发现、高动态应急事件关联挖掘、事件图构建方法等研究，揭示应急事件动态演化机理与关键因果链结构，建立潜在因果识别驱动的持续演化事件图构建算法模型，完成典型态势演化场景下的应急事件图构建模型验证测试，为态势精准分析评估与智能决策提供新途径。

#### 2. 研究内容

(1) 针对持续演化应急事件中多模态信息一致性差与潜在因果关系难以精准识别的难题，开展复杂演化场景下潜

在因果发现方法研究，构建统一的多源异构应急事件时间序列表示，突破基于时序因果框架的多模态应急事件潜在因果发现关键技术，刻画符合应急事件演化机理的潜在因果结构图。

(2) 针对应急事件图动态构建中混淆因素干扰严重、影响事件因果关系准确性的难题，开展去混淆因果推理引导的高动态应急事件关联挖掘方法研究，突破面向潜在因果关系的极小充分混淆变量筛选关键技术，通过系统性检验不变性假设，筛选出需要控制的极小充分混淆变量集合，并研究混淆偏差去除的双重鲁棒因果推理技术，实现应急事件因果关系的精准识别与可信认知。

(3) 针对高动态场景下持续演化事件图构建效率低、准确性难以保障等问题，开展人机协同的持续演化事件图构建框架研究，突破持续演化动态事件图构建关键技术，采用人机协同的方式，实现融合专家研判逻辑与因果模型推理的人机协同事件图动态构建与演化，形成可动态更新的事件图构建算法模型，提升持续演化应急事件图的构建效率以及准确性。

(4) 以数据中台支撑应急事件处置为背景，针对潜在因果识别驱动的持续演化事件图构建方法进行模型应用，验证其在典型场景下对事件因果链条识别与事件演化路径重构的有效性。

### 3. 主要技术指标

(1) 潜在因果发现算法在典型仿真测试集上的因果关系识别  $F1\text{-score} \geq 0.6$ ;

(2) 双重鲁棒因果推理模型在混淆变量干扰场景(混淆强度  $\geq 30\%$ )下的因果效应估计归一化均方根误差(NRMSE)  $\leq 0.15$ 。

### 4. 成果形式

研究报告、算法模型。

### 5. 经费周期

2 年, 20 万。