**东南大学-中国移动研究院联合创新中心**

**知识融合的大小模型协同进化研究及应用**

**联合研发课题指南**

**中国移动通信有限公司研究院**

**2024年03月**

1. **课题基本信息**

|  |  |  |
| --- | --- | --- |
| 课题名称 | | 知识融合的大小模型协同进化研究及应用 |
| 课题所属合作领域 | | 未来业务与信息处理 |
| 课题周期 | | 课题任务合同书签订起——2025年12月31日 |
| 预期成果类型 | | □原型样机 □仿真平台 ☑软件/算法模型 □标准 □白皮书 ☑技术方案或研究报告 ☑专利  ☑论文 □论著 □其他 （填写具体内容） |
| 课题分级 | | □重大课题 □ 重点课题 ☑ 小而美课题 |
| 课题分类 | | ☑技术引领类 □平台验证类 |
| 中国移动课题联系人信息 | 课题接口人 | 黄毅 |
| 联系电话 | 18810325582 |
| 电子信箱 | huangyi@chinamobile.com |

1. **课题研究背景**

近年来，以ChatGPT/ChatGLM等为代表的生成式大模型在语言理解、上下文推理、文本生成等自然语言处理相关任务上表现出令人惊异的能力，这也促进大模型在客服、法律、医疗、气象等行业的应用落地。大模型通常使用自回归方式，首先通过在大量语料上的无监督训练，之后使用有监督训练以及强化学习方式与人类偏好或者行业需求对齐。尽管如此，大模型依然存在生成上的“幻觉”现象，这在一定程度上限制了大模型在实际应用中的可信度，也对安全可控的大模型生成提出了挑战。

在可信度方面，建立外部知识（例如行业知识图谱、概念库、工具类等）与大模型的关联是常用的解决思路，通过大模型内在的通用能力从知识库中检索出相关信息或者选择合适的工具，作为可信的知识来源，为大模型生成提供可追溯的推理路径。在安全可控方面，通过基于人类反馈的强化学习策略可以从训练角度控制模型输出，一定程度上缓解敏感信息的生成；另一方面，在推理中增加控制模块，例如添加带约束的prompt或者能量函数，对输出结果进行限制。

上述研究并未建立起外部知识与大模型之间的内在联系，检索方式存在误差累积现象，安全可控也只能得到缓解。

1. **研究目标及内容**

通过本课题，将实现以下目标：

1. 实现行业知识（数据库、知识图谱、概念库等）与大模型在表征层面有效融合的方式，提出知识增量更新的算法以及大模型中知识修改的策略。

2. 实现推理可追溯、决策透明化的可解释性大模型推理技术，为生成结果提供可靠的知识来源和合理的逻辑路径。

3. 实现以大模型为核心、大小模型协同的多智能体组合策略，并指导生成结果满足预定的各种约束条件。

4. 实现基于多目标优化以及强化学习的多智能体协同进化策略，使得多智能体系统能根据环境变化不断做出演化和改进，从而更好地解决复杂问题。

本课题的主要研究内容包括：

任务1：研究多元外部知识与大模型融合算法；研究可插拔式知识更新算法；

任务2：研究基于思维链及图神经网络的大模型推理算法；

任务3：研究基于隐因子模型及能量函数的多智能体组合策略；

任务4：研究基于多目标优化及强化学习的多智能体协同进化算法。

合作高校负责核心算法和系统的研发和实现。我方参与核心算法研究，负责数据知识以及核心算法和系统的优化验证。

1. **预期成果及验收要求**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **成果名称** | **数量** | **类型** | **交付时间** | **考核指标** | **是否为核心成果** |
| 1 | 面向外部知识与大模型融合的相关技术报告 | 1 | 关键技术研究报告 | 2024年7月 | 完成1篇相关技术研究报告 | 否 |
| 2 | 增量式知识更新算法相关专利 | 1 | 专利 | 2024年10月 | 完成1篇专利撰写并取得专利局申请受理通知书 | 否 |
| 3 | 大模型与外部知识融合/大模型推理算法相关论文 | 2 | 论文 | 2024年12月 | 完成2篇学术论文录用（顶会或顶刊，CCF B/SCI 2区及以上），移动方署名第一作者或通讯作者且署名应排名前二。 | 否 |
| 4 | 多智能体组合及协同进化相关技术报告 | 2 | 关键技术研究报告 | 2025年6月 | 完成2篇相关技术研究报告 | 否 |
| 5 | 基于隐因子模型及能量函数的多智能体组合策略相关的学术论文 | 1 | 论文 | 2025年6月 | 完成1篇学术论文录用（顶会或顶刊，CCF B/SCI 2区及以上），移动方署名第一作者或通讯作者且署名应排名前二。 | 否 |
| 6 | 多智能体协同进化相关专利 | 1 | 专利 | 2025年6月 | 完成1篇专利撰写并取得专利局申请受理通知书 | 否 |
| 7 | 知识融合的大小模型协同进化工作  系统 | 1 | 软件平台或硬件原型 | 2025年6月 | 领域知识生成准确率不低于80%，敏感信息存在率不超过5%，复杂任务完成率不低于90% | 是 |
| 8 | 多智能体组合及协同进化相关技术报告 | 1 | 关键技术研究报告 | 2025年12月 | 完成1篇相关技术研究报告 | 否 |
| 9 | 基于多目标优化及强化学习的多智能体协同进化算法相关学术论文 | 2 | 论文 | 2025年12月 | 完成2篇学术论文录用（顶会或顶刊，CCF B/SCI 2区及以上），移动方署名第一作者或通讯作者且署名应排名前二。 | 否 |

1. **课题执行计划**

课题任务合同书签订起-2024年7月，完成的任务和输出的成果：

面向外部知识与大模型融合的相关技术报告1篇；

知识融合的大小模型协同工作系统原型基础功能实现。

2024年8月-2024年12月，完成的任务和输出的成果：

多元外部知识与大模型融合算法相关学术论文1篇，基于思维链及图神经网络的大模型推理算法学术论文1篇；

增量式知识更新算法相关专利1篇；

知识融合的大小模型协同工作系统原型完成至少两轮迭代优化。

2025年1月-2025年6月，完成的任务和输出的成果：

多智能体组合及协同进化相关技术报告2篇；

基于隐因子模型及能量函数的多智能体组合策略相关的学术论文1篇；

多智能体协同进化相关专利1篇；

知识融合的大小模型协同工作系统满足应用指标要求。

2025年7月-2025年12月，完成的任务和输出的成果：

多智能体组合及协同进化相关技术报告1篇；

基于多目标优化及强化学习的多智能体协同进化算法相关学术论文2篇。

1. **合作团队与配套需求**

1、教授或副教授3名，课题总投入40人月；

2、至少投入博士生140人月、硕士生125人月承担技术研究，课题总投入265人月。

可供课题研究所需的实验室环境，服务器6台等。

1. **知识产权要求**

本课题相关成果的知识产权要求如下：

1. 是否打算就课题成果发表论文：

□我方拟发表 □对方拟发表 ☑共同发表 □均不发表 □不确定

2、本课题所产生的课题成果（不含论文）知识产权归属为：

|  |  |  |
| --- | --- | --- |
| 成果列表 | 知识产权归属 | 成果形式 |
| 成果1：大模型与外部知识融合、多智能体协同相关技术报告 | ☑归中国移动研究院独有  □双方共有 | □专利  □软件  ☑技术秘密  □集成电路布图设计  □其他 |
| 成果2：大模型与外部知识融合、多智能体协同相关专利 | ☑归中国移动研究院独有  □双方共有 | ☑专利  □软件  □技术秘密  □集成电路布图设计  □其他 |
| 成果3：知识融合的大小模型协同工作系统 | ☑归中国移动研究院独有  □双方共有 | □专利  ☑软件  □技术秘密  □集成电路布图设计  □其他 |

**注：知识产权说明以合同约定为准。**

**八、其他说明**

（如有其他内容及说明，请在此补充）