2024年度朝阳区数据要素产业关键共性技术研发课题征集方向

一、数据要素跨域精准登记关键技术研发与行业应用示范

（一）研究内容

随着国家数据基础制度的深化推进，创新与推进符合数据要素特性的数据跨域精准登记技术是发挥数据要素价值、推进数据流通交易的关键，数据要素跨域精准登记是破解数据要素市场发展供给难的首要问题，目前数据登记实践普遍存在缺乏统一标准、统一计量单位、统一语义规范与统一特征指纹等问题，制约了数据登记的效力与效率。

针对数据要素特性，攻关与研发保护提供方数据安全、实现数据数量精准计量、数据特征算法先进、数据语义统一规范的跨域精准登记技术，实现数据要素安全可信登记、统一标识、统一计量、统一语义与统一特征，提高数据要素登记的效率，支持多个登记平台通过协议化方式实现互联互通，防止数据要素重复登记，支持数据要素流通交易、价值评估与入表实践。开展数据要素跨域精准登记的应用示范，支持行业数据要素登记平台的建设与应用，支持构建行业数据要素目录体系，推进行业数据语义开源计划，支持行业间平台互联互通。

（二）考核指标

1.研发数据要素跨域精准登记系统平台，支持行业跨域安全登记并发数≥1000，数据要素计量方法定量精准无异议，应用响应时间小于1秒。

2.研发与建设行业数据要素语义规范化开源平台，支持建设不少于1个行业数据语义规范化开源平台，每个平台语义规范字段数量不少于2万个。

3.支持不少于1个行业应用的数据要素跨域精准登记平台建设，登记数据库数量不少于200个。

4.申请发明专利提交不少于1个，软件著作权不少于3个。

二、基于AI的多源三维空间数据智能处理及分析关键技术研究

（一）研究内容

数据要素是我国经济建设的核心生产要素，数字经济纵深发展促使数据要素规模指数级增长。地理空间数据要素作为基础性、关联性生产要素，已融入生产、分配、流通、消费和社会服务管理等各个环节。在其应用过程中，面临生产与处理效率低、分析预测能力弱等难题。地理信息软件是空间地理数据生产、处理、管理、分析和应用的重要工具，亟需与AI技术深度融合，通过智能化手段进行更全面更快速的数据处理、挖掘与分析，释放海量数据要素价值。

研究三维模型AI自动化构建技术，实现多层级快速建模和城市三维场景自动生成，压缩大场景建模人力和时间成本；实现三维场景语义分割、目标检测与跟踪等能力，提升三维空间的智能化理解与应用水平；研究视频、地图、场景融合技术，开发视频增强与智能分析算法，提高视频图像的质量与可理解性。

（二）考核指标

1.实现LOD1.3、LOD2.0三维模型的AI自动化构建能力，5000个LOD2.0三维模型制作时间不超过20分钟。

2.提供不少于50个空间机器学习和深度学习算子，包括但不限于目标检测、二元分类、地物分类、场景分类、对象提取、变化检测等场景。

3.实现视频空间化并与三维场景融合，支持实时视频接入，支持定点与无人机视频空间化，形成视频AI分析模型不少于5类，包括但不限于目标识别统计、目标轨迹追踪、地理围栏检测等。

4.申请不少于1个发明专利，取得受理通知书；申请不少于3个软件著作权，取得登记证书。

三、混合数据类型的高性能计算框架与优化技术研究

（一）研究内容

在大数据时代，数据类型不再局限于传统的结构化数据，非结构化数据如文本等也日益增多。这些数据的体量庞大，对计算性能的要求极高。AI算法和硬件能力的提升使得处理大规模数据成为可能，但同时也带来了新的挑战。为了更好地进行数据价值挖掘，需要开发能够高效处理混合数据类型的高性能计算框架，并针对不同数据类型优化处理策略。

本研究旨在构建一个能够同时处理结构化和非结构化数据的高性能计算框架，并提出相应的性能优化策略。开发专门的数据处理库，提供统一的API，简化数据加载、转换、分析和存储的过程。应用向量化计算等前沿技术，针对混合数据类型的特定应用场景，如机器学习，开发优化的算法和模型。目标是实现数据处理的高效率、高吞吐量和高可扩展性，同时保持易用性和灵活性。

（二）考核指标

1.设计和实现一个支持混合数据类型的高性能计算框架，该框架能够适应不同数据特性的计算需求，包括结构化数据、文本数据，能够支持AI推理、图计算、SQL查询等算法。

2.研制统一向量化与代码生成两类架构的新型数据处理计算引擎，10TB数据规模的TPC-DS决策支持基准测试的99个查询可以在30分钟内完成。

3.探索高层计算语言在高性能计算环境中的性能瓶颈，并提出相应的优化措施。

4.提交不少于1个专利申请，申请不少于3个软件著作权。

5.成果得到市场验证，服务不少于1家客户的大规模数据处理需求。