

《交通基建大模型评测规范》

（征求意见稿）

编制说明

一、工作简况

（一）任务来源

本文件由中国交通信息科技集团有限公司提出，经中国技术市场协会批准，根据《关于 2024 年度第一批团体标准立项公告》（中技协字〔2024〕23 号）文件，正式列入 2024 年团体标准制修订计划，标准名称为《AI 大模型交通建设图谱架构》。

（二）项目背景

随着新型基础设施智能化转型加速，大模型技术在交通路网规划、施工安全预警、设施智能运维等场景上的应用快速发展。然而，行业面临技术异构性导致的模型兼容性不足、算法黑箱引发的决策可信度存疑、海量基建数据治理标准缺失等核心痛点，在此背景下，编制本规范旨在构建覆盖模型精度、场景泛化性、安全鲁棒性等六大维度的科学评价体系，破解“重研发轻评估”的行业困局，保障交通基建大模型应用的安全可控与可持续发展。

（三）目的意义

制定本文件的目的是为建立行业统一的评估标尺，破解因技术路线差异导致的模型互认难题，为政府采购、企业选型提供科学依据；通过规范化评测牵引技术迭代，重点提升模型在复杂地质条件模拟、多源异构数据处理等交通基建核心场景的工程化能力；从源头构建安全可信的技术应用屏障，明确隐私保护、决策可追溯性等伦理边界，防范因算法缺陷引发的工程质量风险，为

我国交通新基建高质量发展提供标准化保障。

（四）起草单位及起草人名单

本文件起草单位：XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX、XXXXX。

本文件主要起草人：XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX、XXX。

（五）主要起草过程

1. 文本调研

中国交通信息科技集团有限公司于2024年1月启动了文本的调研工作，并于2024年2月完成了相关资料的收集和分析工作。

2. 标准立项

中国交通信息科技集团有限公司向中国技术市场协会标准化委员会提出申请，于2024年3月获得中国技术市场协会标准化工作委员会批准立项。

3. 组建标准起草工作组

2024年4月21日，召开项目启动会。

2024年5月10日，成立了标准起草工作组，并讨论标准调研工作事项。

4. 形成标准草案

2024年7月5日，起草组对资料收集情况进行汇报，并进行线上讨论，开展标准起草前期资料汇总工作。

2024年7月24日，开展组内讨论，确定了标准框架和主要内容。

2024 年 9 月 10 日，完成标准大纲，并组织标准参与单位完成标准大纲评审。

2024 年 12 月 12 日，起草组根据大纲评审意见进行大纲修改，形成标准大纲。

2025 年 2 月~5 月，在确定标准大纲基础上，完成标准草案编制。

2025 年 6 月 11 日，起草组内部组织技术人员对标准草案的技术架构及可实施性进行验证，发现问题并提出改进方向。

2025 年 7 月 10 日，召开标准专家评审会，邀请行业内专家对标准草案进行评审，经专家和起草组讨论，建议标准名修改为《交通基建大模型评测规范》。

5. 形成征求意见稿

2025 年 9 月 17 日，基于各方意见完成标准草案修改，形成标准征求意见稿。

二、确定标准主要内容的论据

（一）编制原则

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》以及《中国技术市场协会团体标准工作程序》的规定起草。

（二）标准主要内容及适用范围

本文件规定了交通基建行业大模型的评测框架、评测能力、评测指标和具体评测方法。

本文件适用于指导测评机构对交通基建行业大模型的多维度能力进行评估、测试等工作，涵盖行业能力、通用能力评测。

（三）确定标准主要内容的论据

首先基于交通建设工程特性，涵盖大模型通用能力和交通基建专用能力，确保评测体系与工程物理场景深度耦合；其次参照《国家人工智能产业综合标准化体系建设指南》技术要求，确立包含准确性、泛化性、可靠性、性能、安全的多维度评价体系；再次依据典型人工智能大模型应用实践，设置评价数据集构建的基本机制；最后参考 GB/T 18905—2002《软件工程产品评价》，规定了评测具体流程。

三、主要试验[或验证]情况分析、技术经济论证、预期经济效果

通过公开数据集测试大模型效果，如以权威的 C-Eval 中文数据集为例，其覆盖 4 大学科大类、52 个小类，涉多领域知识，可测试模型在交通工程、工程管理及相关法规等方面的知识掌握度，考察知识储备与推理能力。

构建行业标准评测集，可参考类似一级建造师考试内容与形式，从知识理解、推理、应用三大维度，按理论知识（60%）、规范应用（30%）、现场实践（10%）的占比，构造交通基建领域标准评测集，精准验证模型行业专业能力。

针对企业内部核心数据难公开的问题，借助数据合成技术形成专业评测集。一方面模拟实际业务场景，测试模型对桥梁设计优化、施工风险预警等业务问题的处理能力；另一方面将合成数据评测结果与少量脱敏私有数据验证结果对比较准，保障评测准确性。

选取典型场景（如路网规划、桥梁健康监测、施工风险预警）

作为测试用例，覆盖不同技术路线（如深度学习、强化学习）、不同规模（十亿级至千亿级参数）的大模型，分别使用公开数据集、企业私有数据及合成数据，围绕标准的科学性、适用性和可操作性展开，编制单位结合实际测试场景和数据，证明其能够有效评估大模型的性能与落地价值。

通过编制交通基建行业大模型评测规范，可以降低行业评测成本，通过标准化评测流程，减少企业重复开发评测工具、重复测试的浪费；可通过权威评测结果，降低市场对模型的信任成本，加速技术采购和商业化落地；可通过评测提前发现模型缺陷，减少因模型误用导致的经济损失。

四、采用国际标准和国外先进标准的程度

本文件为首次自主制定，本文件不涉及国际国外标准的采标情况。

五、重大分歧意见处理经过及依据

本文件在制定过程中未出现重大分歧意见。

六、与现行相关法律、法规及相关标准的协调性

本文件符合国家现行法律、法规和强制性国家标准的要求，参考了 GB/T41867-2022、GB/T42755-2023 等国家标准相关内容要求。

七、知识产权情况说明

本文件不涉及知识产权情况。

八、其他应予说明的事项

无。

《交通基建大模型评测规范》

团体标准起草组

2025 年 9 月 17 日