

团体标准《人工智能 数据集质量评估要求》（征求意见稿）

编制说明

一、工作简况

根据中国电子工业标准化技术协会2023年第一季度第二批团体标准制修订通知，中国电子技术标准化研究院负责《人工智能 数据集质量评估要求》（项目号：CESA-2023-027）标准的制定工作。本标准由中国电子技术标准化研究院提出，由中国电子工业标准化技术协会归口。标准起草单位为中国电子技术标准化研究院、中国科学院软件研究所、中科南京软件技术研究院、上海商汤智能科技有限公司等。

本标准的起草过程简述如下：

2022 年 11 月～2022 年 12 月，中国电子技术标准化研究院牵头组织有关单位，在北京召开了本标准的立项筹备会。会议共有来自中国科学院软件研究所、中科南京软件技术研究院、上海商汤智能科技有限公司等 5 家单位的 14 位专家，对该标准的立项达成共识，共同收集相关资料，形成标准草案大纲。

2023 年 1 月～2023 年 2 月，标准编制组成员多次组织专家进行讨论，对标准草案的技术内容和格式进行修改完善，同时向相关单位和专家发送邮件，征求相关单位的意见和建议。针对搜集到的意见和建议，起草组成员进行了深入细致的研讨，形成了《人工智能 数据集质量评估要求》标准草案。

2023 年 2 月 27 日，中国电子工业标准化技术协会组织有关单位，在北京召开了本标准的立项论证会。本次会议共有来自北京交通大学、北京小米移动软件有限公司、科大讯飞股份有限公司、第四范式（北京）技术有限公司等 4 家单位共 5 位专家对本标准进行了立项论证。专家组一致建议标准发起单位按照专家意见修改完善标准立项建议书，尽快报请协会立项。该标准于 2023 年 3 月正式立项。

会后，标准起草工作组多次召开了标准讨论会，对专家提出的意见、建议进行了认真分析和总结，并进一步修改和完善标准草案，于 2024 年 6 月形成标准征求意见稿。

二、标准编制原则和确定主要内容的论据及解决的主要问题

本标准的制定工作遵循“统一性、协调性、适用性、一致性、规范性”的原则，本着先进性、科学性、合理性和可操作性的原则，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》给出的规则编写。

本标准规定了人工智能领域数据集的质量评估方法，包括通用要求、训练数据集扩展要求、测试数据集扩展要求、数据集安全性、数据集易用性和数据集分享，给出了人工智能领域数据集的通用评估流程。

本标准适用于指导人工智能领域的开发方、用户方以及第三方等相关组织对人工智能领域数据集质量开展评估工作。

三、主要试验[或验证]情况分析

无。

四、知识产权情况说明

本标准不涉及知识产权。

五、产业化情况、推广应用论证和预期达到的经济效果

该标准的编制将为人工智能产品与应用研发过程中的数据集质量评估提出科学客观、可量化评价数据集评价方法及评估流程，有利于规范提升产业数据集质量，全面提升人工智能产业研发效率，促进多行业人工智能产品高质量发展。

六、转化国际标准和国外先进标准情况

无。

七、与现行相关法律、法规、规章及相关标准的协调性

本标准遵循现行法律、法规、规章及标准，与现有标准协调配套。

八、重大分歧意见的处理经过和依据

无。

九、贯彻标准的要求和措施建议

该标准的有效实施将有助于对信息无障碍的推广建设，建议尽快实施。

十、替代或废止现行相关标准的建议

无。

十一、其它应予说明的事项

无。

《人工智能 数据集质量评估要求》

团体标准编制起草组

2024-06-13