



中华人民共和国国家标准化指导性技术文件

GB/Z 236—2026

人工智能 检索增强生成系统评价 指标与方法

Artificial intelligence—Evaluation metrics and methods for retrieval-augmented
generation systems

2026-08-27 发布

国家市场监督管理总局 发布
国家标准化管理委员会

目 次

前言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 检索增强生成系统的评价参考框架 2

6 端到端评价 3

 6.1 答案相关性 3

 6.2 答案正确性 4

 6.3 答案流畅度 4

7 模块化评价 5

 7.1 检索模块评估指标 5

 7.2 生成模块评估指标 7

8 评价过程 7

 8.1 概述 7

 8.2 评价方案确定 7

 8.3 测试数据集准备 7

 8.4 测试环境准备 8

 8.5 测试执行 8

 8.6 测试结果分析 9

附录 A（资料性） 检索增强生成系统的一些其他评价维度 10

 A.1 拒答能力 10

 A.2 信息整合能力 10

 A.3 噪声鲁棒性 10

 A.4 反事实鲁棒性 10

参考文献 12

前 言

本文件为规范类指导性技术文件。

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》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本文件起草单位：中国电子技术标准化研究院、蚂蚁科技集团股份有限公司、北京百度网讯科技有限公司、中国石油天然气股份有限公司塔里木油田分公司、上海文镔信息科技有限公司、中移(杭州)信息技术有限公司、中移(苏州)软件技术有限公司、浪潮软件科技有限公司、浪潮通信信息系统有限公司、厦门渊亭信息科技有限公司、中电信数智科技有限公司、浪潮电子信息产业股份有限公司、中移雄安信息通信科技有限公司、中国电子技术标准化研究院华东分院、浙江大华技术股份有限公司、中移互联网有限公司、上海商汤智能科技有限公司、深信服科技股份有限公司、海信集团控股股份有限公司、江苏赛西科技发展有限公司、中移九天人工智能科技(北京)有限公司、中国移动通信集团有限公司、中兴通讯股份有限公司、上海翰纬信息科技有限公司、北京软件产品质量检测检验中心有限公司、深圳前海微众银行股份有限公司、南京理工大学、北京神舟航天软件技术股份有限公司、北京智芯微电子科技有限公司、中移物联网有限公司、浪潮云信息技术股份公司、山东省计算中心(国家超级计算济南中心)。

本文件主要起草人：徐洋、彭晋、孙曦、马珊珊、张彬琳、李景秋、刘增志、杨其展、马潇、仲凯韬、潘宏博、付涛、李旭东、王珂琛、李照川、梁秉豪、肖红梅、洪万福、朱成忠、陈利明、袁硕之、邸贺亮、郑庆国、严小格、费玥、陈庆洋、韩鹏飞、孔维生、茹宏格、肖夏阳、王和俊、徐金涛、王红伟、田羽慧、张建、董月文、秦日臻、曹汐、张宏伟、史磊、孔昊、高玉翔、戚湧、于连坤、密伟、潘风文、刘勇、郑哲、孔栋、丁源、郑佳佳、逢锦山。

人工智能 检索增强生成系统评价指标与方法

1 范围

本文件提出了检索增强生成系统的测试评价参考框架,给出了端到端的评价指标和模块化评价指标等,并描述了实施评价方法。

本文件适用于检索增强生成系统的开发者、应用者等进行相关产品的开发、验收和应用,也适用于专业评测机构对相关产品和应用效果进行评估评价。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中,注日期的引用文件,仅该日期对应的版本适用于本文件;不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 45288.1—2025 人工智能 大模型 第1部分:通用要求

GB/Z 253—2026 人工智能 检索增强生成通用技术要求

3 术语和定义

GB/T 45288.1—2025 和 GB/Z 253—2026 界定的以及下列术语和定义适用于本文件。

3.1

端到端评价 end-to-end evaluation

评估检索增强生成系统整体性能的活动。

注:端到端评价不单独考察系统中各个独立模块的性能,而是关注系统作为一个整体在完成生成任务时的效果。

该模式更适用于应用方对检索增强生成系统的实现质量进行整体评估、对比或验收。

3.2

模块化评价 module-wise evaluation

对检索增强生成系统中各个独立模块进行独立评估的活动。

注:该模式侧重于考察系统中各个独立模块的实现质量,更适用于开发者和研究者识别系统瓶颈并有针对性地进行开发改进。

3.3

信息点 information point

从问题或答案中抽取出的、能独立表达具体实体、事实、概念、知识、属性或关系的语义单元。

注:在实践过程中,对于信息点的抽取,一般会结合使用工具自动化提取、人工校对和标注的方法。

4 缩略语

下列缩略语适用于本文件。