



中华人民共和国国家标准化指导性技术文件

GB/Z 253—2026

人工智能 检索增强生成通用技术要求

Artificial intelligence—General technical requirements of
retrieval-augmented generation

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 通用技术框架 1

6 知识库要求 2

 6.1 数据接入 2

 6.2 知识获取 2

 6.3 知识切分 3

 6.4 知识增强 3

 6.5 知识索引 4

7 检索增强生成要求 5

 7.1 查询理解 5

 7.2 知识检索 5

 7.3 语义排序 6

 7.4 答案生成 6

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、北京百度网讯科技有限公司、蚂蚁科技集团股份有限公司、百度云智(北京)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、上海文镭信息科技有限公司、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、国网天津市电力公司电缆分公司、江苏金服数字集团人工智能科技有限公司、中国南方电网有限责任公司超高压输电公司、宁波宁帆信息科技有限公司、杭州捷配链数字科技有限公司。

本文件主要起草人：董建、张彬琳、李景秋、刘增志、彭晋、孙曦、徐洋、马珊珊、王金山、袁骁、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、管俊明、李学武、石凡、李君、王宁、彭黔平、杨越。

人工智能 检索增强生成通用技术要求

1 范围

本文件确立了检索增强生成的通用技术框架,规定了检索增强生成、知识库等功能要求。
本文件适用于指导检索增强生成系统的设计和开发。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

检索增强生成 retrieval-augmented generation

从预构建的知识库或数据源中检索相关信息,在特定任务上增强生成式人工智能模型的方法。

3.2

检索增强生成系统 retrieval-augmented generation system

基于检索增强生成技术构建的人工智能系统。

4 缩略语

下列缩略语适用于本文件。

API:应用程序编程接口(Application Programming Interface)

FTP:文件传输协议(File Transfer Protocol)

GIF:图形交换格式(Graphics Interchange Format)

MP3:动态图像专家组第1版标准 音频层第三级(MPEG-1 Audio Layer 3)

MP4:动态图像专家组第4版标准 第14部分(MPEG-4 Part 14)

RAG:检索增强生成(Retrieval-Augmented Generation)

5 通用技术框架

RAG通用技术框架见图1,框架主要包括:

- a) 检索增强生成:通过查询理解、知识检索、语义排序、答案生成等过程,生成与查询问题相关的答案;
- b) 知识库:接入不同来源和异构的数据,进行知识获取和知识切分,抽取实体信息、生成对应的问题、摘要等,来构建查询问题相关的知识库,以便通过向量索引、全文索引等更高效的方式被索引到。