



# 中华人民共和国国家标准

GB/T 41813.2—2022

## 信息技术 智能语音交互测试方法 第2部分：语义理解

Information technology—Intelligent speech interaction testing method—  
Part 2: Semantic understanding

2022-10-12 发布

2023-05-01 实施

国家市场监督管理总局  
国家标准化管理委员会 发布

## 目 次

前言 .....	III
引言 .....	IV
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 概述 .....	1
5 测试准备和执行 .....	2
5.1 测试数据集 .....	2
5.2 测试工具 .....	3
5.3 测试环境 .....	3
5.4 测试执行 .....	3
5.5 结果判定 .....	3
6 功能测试方法 .....	3
6.1 意图理解 .....	3
6.2 命名实体识别 .....	3
6.3 敏感信息辨别 .....	4
6.4 语义拒识 .....	4
6.5 信息检索 .....	4
6.6 文本相似度计算 .....	4
6.7 文本修改 .....	4
6.8 语义修正 .....	4
6.9 自然语言生成 .....	5
6.10 逻辑推理 .....	5
6.11 对话引导 .....	5
6.12 上下文相关的多轮会话 .....	5
7 性能测试方法 .....	5
7.1 语义理解效果 .....	5
7.2 语义理解效率 .....	8
7.3 系统稳定性 .....	9
附录 A (规范性) 主观体验测试 .....	10
A.1 概述 .....	10
A.2 测试项 .....	10
A.3 测试方法 .....	10
参考文献 .....	12

## 前 言

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。

本文件是 GB/T 41813《信息技术 智能语音交互测试方法》的第 2 部分。GB/T 41813 已经发布了以下部分：

- 第 1 部分：语音识别；
- 第 2 部分：语义理解。

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、科大讯飞股份有限公司、小米通讯技术有限公司、华为终端有限公司、深圳市优必选科技股份有限公司、中国电信集团有限公司、思必驰科技股份有限公司、中国科学院自动化研究所、中国医学科学院生物医学工程研究所、哈尔滨工业大学、海信视像科技股份有限公司、马上消费金融股份有限公司、腾讯科技(北京)有限公司、沈阳新松机器人自动化股份有限公司、深圳市人马互动科技有限公司、平安科技(深圳)有限公司、安徽咪鼠科技有限公司、涇丰科技(深圳)有限公司、北京捷通华声科技股份有限公司、北京百度网讯科技有限公司、深圳市北科瑞声科技股份有限公司、阿里云计算有限公司、云从科技集团股份有限公司、上海计算机软件技术开发中心、网易(杭州)网络有限公司、南京云问网络技术有限公司、联想(北京)有限公司、福州数据技术研究院有限公司、国家网络软件产品质量监督检验中心(济南)、华南理工大学、山东省计算中心(国家超级计算济南中心)、神思电子技术股份有限公司、郑州中业科技股份有限公司、中汽数据(天津)有限公司、中国电器科学研究院有限公司、中汽研(天津)汽车工程研究院有限公司、中科极限元(杭州)智能科技股份有限公司、北京爱数智慧科技有限公司。

本文件主要起草人：徐洋、马万钟、吴国纲、朱亚军、贾一君、周立君、宋文林、袁杰、杨震、田定书、钱彦旻、陶建华、花云飞、蒲江波、李海峰、王峰、杨春勇、苏丹、张锋、冯海洪、刘国涛、任军民、陈楠、刑启洲、李笑如、黄石磊、汪淼淼、蔡立志、李军、胡光龙、杨萌、鹿飞、方斌、王岳、井焜、温正棋、李介、张莹、徐向民、孟宪明、高永超、张晴晴。

## 引 言

智能语音交互在智能家居、智能客服、移动终端、车载终端以及智慧教育、智慧医疗、智能办公、服务机器人等诸多领域应用广泛,已成为当前人机交互的重要方式之一。随着智能语音交互的深入,需要对智能语音交互的系统参考框架、基础技术要求、互联网接口要求等进行统一规范,在这方面,国家已制定了支撑智能语音交互系统的基础性国家标准。在此基础上,也需要用统一的测试方法和评价标准来对智能语音交互系统的能力进行评测,为智能语音交互相关的产品和服务提供评测的基础方法和依据。

GB/T 41813《信息技术 智能语音交互测试方法》为 GB/T 36464(所有部分)《信息技术 智能语音交互系统》提供基础通用的测试方法。智能语音交互包括语音识别、语义理解和语音合成三个基本环节,各环节所涉及的测试对象、测试项目、测试环境和测试方法均有所不同。GB/T 41813《信息技术 智能语音交互测试方法》旨在确立和描述适用于智能语音交互各环节的通用测试项和通用测试方法,拟由三个部分构成。

- 第1部分:语音识别。目的在于为智能语音交互应用中的语音识别环节提供通用测试项和通用测试方法。
- 第2部分:语义理解。目的在于为智能语音交互应用中的语义理解环节提供通用测试项和通用测试方法。
- 第3部分:语音合成。目的在于为智能语音交互应用中的语音合成环节提供通用测试项和通用测试方法。

# 信息技术 智能语音交互测试方法

## 第 2 部分：语义理解

### 1 范围

本文件描述了智能语音交互测试中语义理解系统的通用测试项和通用测试方法。

本文件适用于智能语音服务提供商、用户和第三方检测机构对智能语音交互应用的语义理解系统测试的设计和实施。

### 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 36464(所有部分) 信息技术 智能语音交互系统

### 3 术语和定义

GB/T 36464(所有部分)界定的以及下列术语和定义适用于本文件。

#### 3.1

**语义理解 semantic understanding**

使功能单元理解人说话的意图。

[来源：GB/T 36464.1—2020,3.11]

#### 3.2

**命名实体 named entity**

具有特指或唯一含义的指称名字的实体。

#### 3.3

**意图 intention**

语音交互过程中需要系统执行的任务或达成目标。

### 4 概述

智能语音交互测试中的语义理解测试主要包括功能测试与性能测试，具体如下：

——功能测试用于检查被测系统是否提供了语义理解相关的各项功能，包括：意图理解、命名实体识别、敏感信息识别、语义拒识、信息检索、文本相似度计算、文本修改、语义修正、自然语言生成、逻辑推理、对话引导和上下文相关的多轮对话；

——性能测试用于检查被测系统中语义理解相关的各项性能，包括：语义理解效果、语义理解效率和系统稳定性。

测试过程中，应依据被测系统技术规范进行在线/离线的功能和性能测试。可使用主观体验的方法进行测试，主观体验测试应按照附录 A 描述的测试项和测试方法执行。本文件所列测试项可用于对