

国家标准

《信息技术 智能语音交互测试 第2部分：语义理解》

（征求意见稿）编制说明

一、工作简况

根据国家标准化管理委员会 2020 年下达的国家标准制修订计划，国家标准《信息技术 智能语音交互测试 第2部分：语义理解》（计划号：20194129-T-469），由全国信息技术标准化技术委员会提出并归口。该标准由中国电子技术标准化研究院、科大讯飞股份有限公司负责起草。

标准计划下达后，标准起草组进行了在全国信标委组织领导下，按计划开展本标准制定工作，具体内容如下：

2020 年 2 月，全国信标委用户界面分技术委员会向社会各界征集联合起草单位，成立标准起草组。起草组基于前期智能语音交互测试项和测试方法的标准化研究以及验证工作，包括：起草标准草案、开发智能语音交互测试工具、搭建技术验证平台以及验证标准条款等，制定标准范围和内容框架；

2020 年 3 月，起草组召开技术讨论会，正式开始起草标准工作组草案。起草组在对标准的关键技术进行讨论和修改后，形成标准工作组草案；

2020 年 4 月，起草组对本标准的测试项和测试方法逐一开展测试并分析，对标准工作组草案进行第一次修改；

2020 年 5 月，起草组在内部开展了内容评议，共征集到修订意见 89 项，起草组整理和讨论了各单位的反馈内容；

2020 年 6 月，起草组落实各项意见与建议，对标准工作组草案进行第二次修改。

2020 年 7 月，起草组将标准草案提交全国信息技术标准化委员会进行开题评审，经与专家讨论，将标准名称修改为《信息技术 智能语音交互测试方法 第2部分：语义理解》。

2020 年 8 月，起草组对标准草案展开讨论，并进行技术内容制定和完善，对标准工作组草案进行第三次修改。

2020 年 9 月，起草组将标准草案提交全国信息技术标准化技术委员会进行

内审，经与专家讨论，将标准结构进行编辑性调整，并根据起草组内意见对标准草案进行进一步完善，形成了标准征求意见稿。

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

GB/T XXXXX《信息技术 智能语音交互测试方法》为GB/T 36464《信息技术 智能语音交互系统》提供基础性测试方法。本标准为GB/T XXXXX《信息技术 智能语音交互测试方法》的第1部分。GB/T XXXXX包括以下部分：

——第1部分：语音识别；

——第2部分：语义理解；

——第3部分：语音合成。

本标准结构如下：

● 前言

介绍标准起草的依据，与其它标准的关系，标准的基本信息，包括标准结构、提出和归口单位、起草单位与起草人。本标准是GB/T XXXXX《信息技术 智能语音交互测试》的第2部分。

● 引言

介绍编制标准的原因、目的和分部分的原因等事项说明。

● 范围

本文件描述了智能语音交互测试中语义理解(子)系统的测试项和测试方法。本文件适用于智能语音交互测试中语义理解(子)系统测试的设计和实施。

● 术语、定义和缩略语

定义本标准正文中出现的、需要解释的术语。

● 测试项

按照语义理解相关的功能和性能，描述智能语音交互语义理解测试的测试项和测试指标。

● 测试方法

按照测试数据集、测试工具和设备、测试环境、测试执行和结果判定，给出功能测试和性能测试的测试方法。

本标准是GB/T XXXXX《信息技术 智能语音交互测试方法》的第2部分。本部分所涉及的内容为已发布实施的GB/T 36464《信息技术 智能语音交互系统》

提供支撑。

本标准将为智能语音交互系统在智能家居、智能客服、移动终端、车载终端等应用领域和场景提供基础性、通用性的测试方法，统一和规范国内智能语音交互在语义理解测试中的术语、要求、方法等方面的内涵，减少各标准之间内容的重复和冲突；也将为智能语音交互系统在其他应用领域和方向上的标准化工作提供指导，使智能语音交互技术在不同领域应用过程中可检测，提升该领域国家、行业、地方和团体标准之间的协调性和一致性。

鉴于国内该领域发展与研究的现状和趋势，以及国际标准化进展，本标准为自主研发的国家标准，具体方法如下：

- 充分调研行业发展现状，坚持以行业内技术和应用实际为基础，又兼顾未来的需求的原则，确保促进技术和应用的高质量发展；
- 充分调研使用群体需求，坚持用户使用为先导，做好技术、产品和服务的通用要求，规范市场准入，保障用户权益。

三、主要试验[或验证]情况分析

本标准规范的智能语音交互语义理解测试已应用于智能家居、智能客服、移动终端、车载终端等多个应用场景，并已作为新一代人工智能技术评测能力之一，应用于智能语音国家人工智能开放创新平台，通过在线测试、离线测试等多种形式，对各种类型的智能语音交互产品和服务进行了测试。

1、测试项确认

以人工智能、语音技术和人机交互等理论为基础，以智能语音交互系统技术为依据，以实际产品和服务开发为验证平台，确定智能语音交互语义理解测试包括功能测试项和性能测试项。其中，功能测试项包括意图理解、意图分类、命名实体识别、语义槽抽取、敏感信息辨别、语义拒识、信息检索、文本修改、语义修正、自然语言生成、逻辑推理、对话引导等；性能测试项包括理解效果、理解效率和系统稳定性等。各测试项包括其对应的测试指标和计算方法。

2、测试方法确认

以智能语音交互系统产品和服务为基础，调研不同产品和服务在设计、开发和测试过程中积累的实际经验，并充分考虑终端用户实际可感知的评价维度，抽取国内主流的智能语音技术产品和服务提供商的共有测试指标，确定智能语音交互语义理解测试的测试方法，包括测试数据集、工具设备和测试流程。同时，针

对 GB/T 36464《信息技术 智能语音交互系统》中出现的技术项和要求进行比对，对本标准内容进行进一步完善，以保证本标准对已有国家标准的支撑作用以及可操作和可复用性。

四、知识产权情况说明

本标准不涉及专利和知识产权问题。

五、采用国际标准和国外先进标准情况

本标准为自主研发的国家标准，未采用国际标准和国外先进标准。

六、与现行相关法律、法规、规章及相关标准的协调性

本标准符合我国有关的现行法律、法规和相关强制性标准的相关规定，并与相关标准保持协调一致。

七、重大分歧意见的处理经过和依据

无。

八、标准性质的建议

鉴于本标准的内容，建议该标准作为推荐性国家标准发布。

九、贯彻标准的要求和措施建议

为更好地发挥标准实施效力，建议采取以下措施推动本标准的贯彻实施：

1、 政府积极推动：标准发布后，由国家市场监管总局、国家工信部指导，由全国信息技术标准化技术委员会牵头，组织召开标准发布通报会，并通过官方媒体在相关产业链和应用领域加强宣传，提升标准的宣传权威性和受众针对性；

2、 制定配套实施政策：政府通过制定配套实施政策，如在招投标、市场准入、认证服务、企业对标达标等方面出台相关政策，加强标准的实施力度，引导和提升相关企业和单位应用标准积极性；

3、 加强标准检测认证一体化：鼓励第三方依据和采用标准进行产品试验、检测和认证，开展产品国家标准符合性测试，同时推动建立政府和市场的多方标准质量采信机制，不断提升标准实施的效力。

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。

国家标准《信息技术 智能语音交互测试 第2部分：语义理解》

起草工作组

2020-10