

# 人工智能语音应用开发 职业技能等级标准

(2021年1.0版)

科大讯飞股份有限公司 制定  
2021年4月 发布

# 目次

前 言.....	II
1 范围.....	1
2 规范性引用文件.....	1
3 术语与定义.....	2
4 适用院校专业.....	3
5 面向职业岗位（群）.....	4
6 职业技能要求.....	4
参考文献.....	17

## 前 言

本标准按照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。

本标准起草单位：科大讯飞股份有限公司、国家工业信息安全发展研究中心、安徽信息工程学院、深圳职业技术学院、常州信息职业技术学院、重庆电子工程职业学院、深圳信息职业技术学院、淄博职业学院、武汉职业技术学院、襄阳职业技术学院、重庆工商职业学院、广东水利电力职业技术学院、湖北职业技术学院、贵州交通职业技术学院、徐州工业职业技术学院、广西经贸职业技术学院、广西理工职业技术学院、广东科学职业技术学院、安徽林业职业技术学院、广东理工职业学院、苏州信息职业技术学院、云南交通职业技术学院、湖南软件职业学院、苏州市职业大学、重庆财经学院、广西科技师范学院、广东财经大学、广东技术师范大学、广东开放大学、广西大学行健文理学院、广西外国语学院、南宁师范大学、南宁学院、四川大学锦江学院、淮阴工学院。

本标准主要起草人：陈涛、周佳峰、何小龙、吴华安、李栋学、莫少林、吴有富、蔡铁、武春岭、胡方霞、魏本征、贺敏伟、曾文权、孙宾、张卫东、杨勇、钱银中、凌明胜、刘小华、李粤平、王宝成、胡昌杰、肖政宏、李向前、吴砥、刘晓、桂诚、丁德成、胡江院、崔小蕾、马季、雷大正、殷振华、张涛、陈小贝。

声明：本标准的知识产权归属于科大讯飞股份有限公司，未经科大讯飞股份有限公司同意，不得印刷、销售。



## 1 范围

本标准规定了人工智能语音应用开发职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于人工智能语音应用开发职业技能培训、考核与评价，相关用人单位的人员聘用、培训与考核可参照使用。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本适用于本文件。

国际标准: W3C Recommendation, Speech Synthesis Markup Language (SSML)  
Version 1.1

国家、行业有关标准如下:

GB/T 36464 《信息技术 智能语音交互系统》

GB/T 21023—2007 中文语音识别系统通用技术规范

GB/T 21024—2007 中文语音合成系统通用技术规范

GB/T 34083—2017 中文语音识别互联网服务接口规范

GB/T 34145—2017 中文语音合成互联网服务接口规范

SJ/T 11380—2008 自动声纹识别（说话人识别）技术规范

### 3 术语与定义

国家、行业标准界定的以及下列术语和定义适用于本文件。

#### 3.1. 语音交互 (speech interaction)

人类和功能单元之间通过语音进行的信息传递和交流活动。

#### 3.2. 语音交互系统 (speech interaction system)

由功能单元（或其组合）、数据资源等组成的能够实现与人类之间进行交互的语音系统。

#### 3.3. 在线 (online)

处于跟互联网相联的状态。

#### 3.4. 离线 (offline)

处于跟互联网连接断开的状态。

#### 3.5. 语音识别 (speech recognition)

将人类的声音信号转化为文字或者指令的过程。

【GB/T 21023-2007，定义 3.1】

#### 3.6. 语音合成 (speech synthesis)

通过机械的、电子的方法合成人类语音的过程。

【GB/T 21024-2007，定义 3.1】

#### 3.7. 语音听写 (speech auto transformation)

是基于自然语言处理，将自然语言音频转换为文本输出的技术。

### 3.8. 语音唤醒 (voice wakeup)

通过辨别输入的音频中特定的词语，返回被命中（唤醒）结果，应用通过回调的结果，进行下一步的处理，如点亮屏幕，或与用户进行语音交互等。

### 3.9. 语音评测 (speech evaluator)

在计算机辅助语言学习、考试等系统中，对发音人的发音水平进行分析和评判的过程。

### 3.10. 声纹识别 (voiceprint recognition)

根据待识别语音的声纹特征识别该段语音所对应的说话人的过程。

### 3.11. 语音增强 (speech enhancement)

指当语音信号被各种噪声干扰，甚至淹没后，从噪声背景中提取有效语音信号，抑制、降低噪声干扰的技术。

## 4 适用院校专业

中等职业学校：计算机应用、软件与信息服务、移动应用技术与服务、数字媒体技术应用、电子与信息技术、数字媒体技术应用、物联网技术应用等专业。

高等职业学校：人工智能技术服务、计算机应用技术、软件技术、软件与信息服务、计算机应用技术、计算机系统与维护、智能产品开发、智能控制技术、数字媒体应用技术、移动应用开发、移动互联应用技术、物联网应用技术、物联网工程技术等专业。

应用型本科学校：人工智能、计算机应用工程、软件工程、计算机科学与技术、电子与计算机工程、计算机应用工程、智能科学与技术、智能控制技术、信息工程、数字媒体技术、物联网工程等专业。

## 5 面向职业岗位（群）

主要面向包含智能语音软件开发、智能硬件开发、数据服务、智能语音技术应用等 IT 企业，从事语音数据收集、预处理、标注，智能语音 Python 程序开发，Web 应用开发，机器学习算法集成，深度学习模型训练、调优，人工智能语音算法应用等工作。面向的主要岗位包括等语音数据标注员、语音数据预处理工程师、人工智能算法工程师、智能语音系统开发工程师等。

## 6 职业技能要求

### 6.1 职业技能等级划分

人工智能语音应用开发职业技能等级分为三个等级：初级、中级、高级，依次递进，高级别涵盖低级别技能要求。

**【人工智能语音应用开发】(初级):** 主要面向软件开发、智能硬件开发、数据服务、人工智能语音技术应用等类型企业，从事语音数据收集、数据预处理、数据标注等工作，能遵循数据标注规范及要求，具有能使用智能语音数据标注平台进行语音识别数据标注、语音合成数据标注、声纹识别数据标注等人工智能经典场景数据标注能力；具有使用 Python 语句完成基础软件制作的能力；具有使用 NumPy 进行矩阵运算、数据标准化等运算的能力；具有使用 Pandas 进行 csv 数据的加载、数据筛选、数据合并、NumPy 数据转换等数据处理能力。



**【人工智能语音应用开发】(中级):** 主要面向软件开发、智能硬件开发、数据服务、智能语音技术应用等类型企业，从事人工智能语音 Python 程序开发、Web 应用开发、机器学习算法集成、深度学习模型训练、人工智能算法应用等工作，具有使用 Python、Web 应用程序等工具进行智能语音应用开发能力；具有在应用软件、移动端应用、Web 前端应用中使用 scikit-learn 进行机器学习基础模型应用的能力。

**【人工智能语音应用开发】(高级):** 主要面向软件开发、智能硬件开发、智能语音技术应用等类型企业，从事机器学习算法集成、深度学习模型训练、深度学习模型调优、深度学习模型应用等工作，具有简单智能语音算法实现、对智能语音 SDK 进行二次封装及构建高性能智能语音应用程序的能力。

## 6.2 职业技能等级标准描述

表 1 人工智能语音应用开发职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1. 数据采集	1.1 程序设计与开发	<p>1.1.1 能够搭建 Python 基本运行环境，能够熟悉 Python 脚本和控制台两种模式。</p> <p>1.1.2 能够使用 Python 列表、元组、字典、字符串等基本数据结构实现数据的存储及其他操作。</p> <p>1.1.3 能够使用 Python 条件控制和循环控制实现逻辑处理功能。</p> <p>1.1.4 能够使用 Python 模块调用和自定义模块的方法实现模块化的设计形式。</p> <p>1.1.5 使用 conda 或 pip 包管理工具，能够在命令行配置和管理需要的 Python</p>

		库。
	1.2 语音数据采集工具应用	<p>1.2.1 具备使用数据采集工具进行语音数据采集的能力。</p> <p>1.2.2 能够根据语音识别模型要求使用网络爬虫获取语音识别训练数据。</p> <p>1.2.3 能够使用数据采集工具模拟浏览器，完成数据的正常采集。</p> <p>1.2.4 能够使用数据采集工具完成对App端数据的采集。</p>
	1.3 数据采集与存储	<p>1.3.1 具备使用数据采集工具或自定义脚本的方式进行数据采集的能力。</p> <p>1.3.2 能够在数据采集工具中集成数据清理模块，完成数据采集到清理的流程。</p> <p>1.3.3 能够在数据采集工具中集成数据存储模块，完成数据采集到数据存储的流程。</p> <p>1.3.4 能够结合语音应用对行业数据的要求完成语音数据的采集。</p>
2. 数据预处理与展示	2.1 数据预处理(Numpy)	<p>2.1.1 能够使用NumPy 的基本数据结构多维数组(ndarray)和多维矩阵(matrix)存储数据。</p> <p>2.1.2 使用数组的切片和高级索引，能够灵活的取出数组中的部分数据。</p> <p>2.1.3 能够灵活地对数组以整体进行代数数值运算。</p> <p>2.1.4 能够实现数组的改变形状、转置、修改维度、连接、分割等操作。</p> <p>2.1.5 能够熟练地计算数组中的最大值、最小值、均值、方差等基本统计量。</p> <p>2.1.6 能够灵活地对数组按不同的维度排序。</p> <p>2.1.7 能够使用NumPy 拷贝与视图的机</p>

		<p>制对二者在不同的函数调用中清晰地区分。</p> <p>2.1.8 能够灵活地调用 NumPy 的矩阵和线性代数函数。</p> <p>2.1.9 能够读取和保存 NumPy 数组。</p> <p>2.1.10 能够熟练对语音识别、语音合成、语音评测、声纹识别等语音数据进行归一化、去异常数据等清洗和预处理操作。</p>
	2.2 数据预处理(pandas)	<p>2.2.1 理解 Pandas 基本数据结构：Series、DataFrame、Panel。</p> <p>2.2.2 能够利用 pandas 提供的描述性统计函数进行数据地初步分析。</p> <p>2.2.3 能够使用函数应用的功能实现对表格中元素的处理。</p> <p>2.2.4 能够实现对数据元素的高效获取。</p> <p>2.2.5 能够从各个角度对数据进行较为复杂的统计分析。</p> <p>2.2.6 能够有效地在内存中将表格数据合并和级联。</p>
	2.3 数据可视化	<p>2.3.1 熟练使用 Excel、Power BI 等可视化工具，并能使用可视化工具进行数据可视化。</p> <p>2.3.2 能够熟练地绘制折线图、饼图、直方图、散点图、盒图等基本图形进行数据展示。</p> <p>2.3.3 能够将 Pandas 数据通过 Seaborn、Matplotlib 类库进行展示。</p> <p>2.3.4 能够通过数据展示工具或数据展示脚本展示语音数据表达的频谱。</p>
3. 语音数据采集与标注	3.1 数据标注工具应用	3.1.1 能够根据数据标注工具的要求完成不同的任务。

		<p>3.1.2 能够在实际工作中熟练应用语音标注工具或产品进行数据标注。</p> <p>3.1.3 能使用数据标注工具结合语音识别应用要求对语音数据进行标注。</p> <p>3.1.4 能够基于标注平台的规则完成系统管理员、应用管理员和标注员的规则完成不同的任务。</p>
	3.2 数据标注	<p>3.2.1 能够理解不同的模型对数据的标注要求。</p> <p>3.2.2 能够根据不同的算法模型对数据的要求完成模型数据的标注。</p> <p>3.2.3 能够使用工具将数据从数据平台导入标注平台，并进行分类。</p> <p>3.2.4 能够对标注的数据进行检测和测试。</p>
	3.3 语音数据标注	<p>3.3.1 能使用标注工具对语音识别数据进行标注，并制作语音识别训练数据集。</p> <p>3.3.2 能够使用语音标注工具，按照标注规范对音频内容人工转写成对应文本。</p> <p>3.3.3 能够熟练应用不同场景下的智能语音标注工具或产品，并能对使用过程中产生的数据进行收集和分析，为工具或产品的优化提供依据和支撑。</p> <p>3.3.4 能够熟练使用标注工具对语音识别、语音合成、语音评测、声纹识别等语音数据集进行标注。</p>

表 2 人工智能语音应用开发职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
------	------	--------

1. 程序设计与开发	1.1 程序优化	<p>1.1.1 熟练掌握面向对象思维，并能够基于 Python 语言进行面向对象程序设计。</p> <p>1.1.2 能够遵循项目设计，完成项目编码。</p> <p>1.1.3 能够使用模块、类等知识对 Python 源码进行封装。</p> <p>1.1.4 能够使用正则表达式匹配、筛选各种形式的数据库。</p> <p>1.1.5 能够使用 Python 进行文件读写。</p>
	1.2 数据存储	<p>1.2.1 掌握 T-SQL 语句，能够使用 T-SQL 语句进行数据库、数据表的创建。</p> <p>1.2.2 能够使用 Insert、Update 等语句完成数据的增、删、改、查操作。</p> <p>1.2.3 能够使用 Python 数据库接口实现数据库的增、删、改、查操作。</p> <p>1.2.4 能够使用 ORM 工具进行面向对象的数据结构设计和存储。</p>

	1.3 Web 编程	<p>1.3.1 了解 TCP/IP 等通信协议。</p> <p>1.3.2 能够使用 Socket 结合 TCP 协议实现简单的局域网通信。</p> <p>1.3.3 能够实现简单的网络编程任务。</p> <p>1.3.4 能够使用 GUI 编程方法编写简单的图形交互接口。</p>
2. 算法设计与实现	2.1 数据处理	<p>2.1.1 能够使用 scikit-learn 实现数据集的加载。</p> <p>2.1.2 能够使用 scikit-learn 实现数据的预处理。</p> <p>2.1.3 能够使用 scikit-learn 实现特征的提取。</p> <p>2.1.4 能够对处理之后的数据进行可视化。</p>

	2.2 算法设计与选择	<p>2.2.1 熟悉经典的机器学习算法，并能够根据不同的问题选择不同的机器学习算法。</p> <p>2.2.2 能够使用特征选择算法对训练数据进行特征选择。</p> <p>2.2.3 能够使用模型评估算法对机器学习模型进行评估。</p> <p>2.2.4 能够通过工具对比不同算法模型。</p>
	2.3 算法实现	<p>2.3.1 能够使用 scikit-learn 实现模型验证和评估。</p> <p>2.3.2 能够使用 scikit-learn 实现模型训练和预测。</p> <p>2.3.3 能够使用 scikit-learn 实现模型持久化，能够将训练好的模型参数保存，并能够在原有基础上再次训练。</p> <p>2.3.3 能够使用 scikit-learn 实现基本的机器学习算法，如支持向量机、决策树、朴素贝叶斯、聚类分析、神经网络等常见模型，掌握集成方法如 Bagging、AdaBoost、随机森林、梯度提升树等。</p>

3 智能语音应用开发	3.1 综合应用开发	<p>3.1.1 能够通过调优使语言和声学模型达到特定的准确率/召回率指标；熟悉常用信号处理、通用机器学习和深度学习算法的原理及不同算法间的差异及相关适用场合等。</p> <p>3.1.2 具备一定的语音翻译、语音控制、语音转录、情感识别及声纹识别等语音识别应用的开发能力。</p> <p>3.1.3 熟悉 MVP、MVC 等 Application 层开发、调优框架，能基于智能语音 SDK 开发高性能智能语音 Application。</p> <p>3.1.4 具备面向对象设计思维，针对不同的开发平台能使用常见的设计模式解决应用问题。</p>
	3.2 框架应用（Flask）	<p>3.2.1 能够灵活地构建 URL。</p> <p>3.2.2 能够使用 Flask 中模板引擎建立 flask 框架。</p> <p>3.2.3 能够使用消息闪现（Flash()）传递数据。</p> <p>3.2.4 能够实现文件上传的功能。</p>
	3.3 应用发布	<p>3.3.1 熟悉 RESTful 标准，能够使用 Flask 框架构建 RESTful 接口。</p> <p>3.3.2 能够基于 Flask 框架开发网络请求验证规则，设计验证算法。</p> <p>3.3.3 能够使用 Scikit-Learn 实现并保存算法模型，并使用 Flask 构建机器学习服务程序。</p> <p>3.3.4 能够部署 Flask 应用程序。</p>

表 3 人工智能语音应用开发职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
1. 智能语音算法	1.1 语音合成	1.1.1 熟悉传统语音合成与识别流程。



实现（必须掌握 4 种职业技能种的 2 种）	与语音识别	<p>1.1.2 熟悉新一代语音合成技术，能使用 TensorFlow 等深度学习框架实现 seq2seq 模型，并能使用 Wavenet 等算法将频谱转换为语音。</p> <p>1.1.3 熟悉语音识别框架，熟练掌握常见的语言模型框架（N-gram, RNN 等），掌握声学模型框架（HMM LSTM/CNN 框架），了解新一代语音识别框架，掌握 CTC、Encoder-Decoder 等技术框架。</p>
	1.2 语音增强	<p>1.2.1 熟悉语音增强，掌握从噪声背景中提取有用的语音信号，抑制、降低噪声干扰的技术。</p> <p>1.2.2 掌握传统单通道语音增强技术，掌握深度学习单通道语音增强技术。</p> <p>1.2.3 掌握麦克风阵列语音增强技术与深度学习麦克风阵列多通道增强技术。</p>
	1.3 语音测评	<p>1.3.1 熟悉语音测评的流程，熟悉语音测评包含内容（朗读测评与口头表述测评）。</p> <p>1.3.2 熟练掌握语音测试核心技术：发音检错，朗读评分，自由表达评分。</p>
	1.4 声纹识别	<p>1.4.1 了解声纹识别应用，了解生理与行为的特征，熟悉声纹识别的技术原理。</p> <p>1.4.2 了解新一代声纹识别技术，掌握基于 CNN 的声纹识别技术路线，熟悉其它生物识别技术（指纹识别，虹膜识别）。</p>
2. 程序设计与开发与部署	2.1 环境搭建	<p>2.1.1 能够搭建 Python 基本运行环境，能够熟悉 Python 脚本和控制台两种模式。</p> <p>2.1.2 使用 conda 或 pip 包管理工具，能够在命令行配置和管理需要的 Python 库。</p> <p>2.1.3 能够使用 Anaconda 命令行模式创建深度学习虚拟环境，安装 TensorFlow、</p>

		<p>Scikit-Learn 等机器学习库。</p> <p>2.1.4 能够在 PyCharm、Visual Studio Code 中配置和使用 Anaconda 虚拟环境。</p>
	2.2 程序设计与开发	<p>2.2.1 能够使用 Python 中的列表、元组、字典、字符串等基本数据结构实现数据的存储及其他操作。</p> <p>2.2.2 能够使用 Python 条件控制和循环控制实现逻辑处理功能。</p> <p>2.2.3 能够使用 Python 模块调用和自定义模块的方法实现模块化的设计形式。</p> <p>2.2.4 能够熟练使用 Python 面向对象思想设计和实现应用程序。</p> <p>2.2.5 能够使用 Python 正则表达式匹配、筛选各种形式的数据。</p> <p>2.2.6 能够使用 Python 进行文件读写。</p> <p>2.2.7 能够使用 Python 数据库访问实现对数据库的增删改查操作。</p> <p>2.2.8 能够实现简单的网络编程任务。</p> <p>2.2.9 能够使用 GUI 编程方法编写简单的图形交互接口。</p>
	2.3 算法实现	<p>2.3.1 能够搭建 TensorFlow 环境，能够使用 CUDA 库调用显卡加速 TensorFlow 的计算。</p> <p>2.3.2 能够用代码构建深度学习算法的计算图。</p> <p>2.3.3 能够用 TensorFlow 提供的优化算法，实现对网络参数的优化。</p> <p>2.3.4 能够实现断点续训和对预训练参数的加载。</p> <p>2.3.5 能够利用 Tensorboard 可视化工具，可视化训练过程和查看网络构建情况。</p>

3 智能语音应用开发与部署	3.1 综合应用开发	<p>3.1.1 能够通过调优使语言和声学模型达到特定的准确率/召回率指标；熟悉常用信号处理、通用机器学习和深度学习算法的原理及不同算法间的差异及相关适用场合等。</p> <p>3.1.2 具备基于智能语音 SDK 构建语音翻译、语音控制、语音转录、情感识别及声纹识别等语音识别应用的构建能力。</p> <p>3.1.3 熟悉 MVP、MVC 等 Application 层开发、调优框架，能基于智能语音 SDK 构建高性能智能语音 Application。</p> <p>3.1.4 具备根据业务需求，对智能语音 SDK 进行二次封装的能力。</p> <p>3.1.5 熟悉网络通信机制，能基于 Socket 完成简单的网络通信应用程序。</p> <p>3.1.6 对 Http、Https、TCP/IP 等通信协议有深入的理解。</p> <p>3.1.7 具备 Application 的性能优化能力，能够从加载、渲染、执行、查询等角度对 Application 进行性能优化。</p>
	3.2 框架应用 (Flask)	<p>3.2.1 能够灵活地构建 URL。</p> <p>3.2.2 能够使用 Flask 中模板引擎建立 flask 框架。</p> <p>3.2.3 能够使用消息闪现 (Flash()) 传递数据。</p> <p>3.2.4 能够实现文件上传的功能。</p>
	3.3 应用发布	<p>3.3.1 熟悉 RESTful 标准，能够使用 Flask 框架构建 RESTful 接口。</p> <p>3.3.2 能够基于 Flask 框架开发网络请求验证规则，设计验证算法。</p> <p>3.3.3 能够使用 Scikit-Learn 实现并保存算法模型，并使用 Flask 构建机器学习</p>

		服务程序。 3.3.4 能够部署 Flask 应用程序。
--	--	---------------------------------

## 参考文献

- [1] 中华人民共和国教育部,《中等职业学校专业目录》(2010年修订)[M].北京:高等教育出版社,2010-11.
- [2] 中华人民共和国教育部,《中等职业学校专业目录》增补专业(2019年)[Z].
- [3] 中华人民共和国教育部,《普通高等学校高等职业教育(专科)专业目录》(2015年)[Z].
- [4] 中华人民共和国教育部,《普通高等学校高等职业教育(专科)专业目录》2019年增补专业[Z].
- [5] 中华人民共和国教育部,《普通高等学校(本科)专业目录》(2012年)[Z].
- [6] W3C Recommendation, Speech Synthesis Markup Language (SSML) Version 1.1
- [7] GB/T 36464 《信息技术 智能语音交互系统》
- [8] GB/T 21023—2007 中文语音识别系统通用技术规范
- [9] GB/T 21024—2007 中文语音合成系统通用技术规范
- [10] GB/T 34083—2017 中文语音识别互联网服务接口规范
- [11] GB/T 34145—2017 中文语音合成互联网服务接口规范
- [12] SJ/T 11380—2008 自动声纹识别(说话人识别)技术规范