

团 体 标 准

T/ CBA 220—2021

远程银行人工智能客服评价指标规范

Specification for evaluating indicators of AI customer service of the remote banking

2021 - 09 - 16 发布

2021 - 09 - 16 实施



中国银行业协会 发布

目 次

前 言.....	II
引 言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 评价指标的范围、方法及内容.....	2
4.1 评价指标的范围.....	2
4.2 评价指标的方法.....	2
4.2.1 抽样法.....	2
4.2.2 报表法.....	3
4.3 评价指标的内容.....	3
4.3.1 语音识别准确率.....	3
4.3.2 问题识别率.....	4
4.3.3 意图理解准确率.....	4
4.3.4 交互准确率.....	5
4.3.5 问题解决率.....	5
4.3.6 智能分流率.....	5
5 综合评价指标权重.....	6
参 考 文 献.....	7

前 言

中国银行业协会(China Banking Association, CBA)成立于2000年5月,是经中国人民银行和民政部批准成立,并在民政部登记注册的全国性非营利社会团体,是中国银行业自律组织。2003年中国银监会成立后,中国银行业协会主管单位由中国人民银行变更为中国银监会。2018年3月,中国银行保险监督管理委员会成立后,中国银行业协会主管单位由中国银监会变更为中国银行保险监督管理委员会。凡经业务主管单位批准设立的、具有独立法人资格的银行业金融机构(含在华外资银行业金融机构)和经相关监管机构批准、具有独立法人资格、在民政部门登记注册的各省(自治区、直辖市、计划单列市)银行业协会以及相关监管机构批准设立,具有独立法人资格的依法与银行业金融机构开展相关业务合作的其他类型金融机构,以及银行业专业服务机构均能申请加入中国银行业协会成为会员单位。

中国银行业协会日常办事机构为秘书处。秘书处设秘书长1名,副秘书长若干名。根据工作需要,中国银行业协会设立32个专业委员会,其中银行业产品和服务标准化专业委员会旨在开展银行业产品和服务标准化工作,包括制定和发布银行业的产品和服务标准,积极参与制定国家标准、行业规划,参与制定有关政策和法律法规,不断提高银行业产品和服务质量。

本文件按照T/CBA 1—2021《中国银行业协会团体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这些专利的责任。

本文件由中国银行业协会远程银行专业委员会提出。

本文件由中国银行业协会产品与服务标准化专业委员会归口。

本标准主要起草单位:中国银行业协会、中国建设银行股份有限公司、广发银行股份有限公司、中国工商银行股份有限公司、交通银行股份有限公司、中国邮政储蓄银行股份有限公司、中信银行股份有限公司、上海浦东发展银行股份有限公司、中国民生银行股份有限公司、华夏银行股份有限公司、平安银行股份有限公司、兴业银行股份有限公司、北京银行股份有限公司、杭州银行股份有限公司。

本标准主要起草人员:潘光伟、刘峰、高峰、张庚、伊贵英、李宽、成星、任旭华、赵成刚、胡冬梅、刘鑫、吕林、仲峻锋、陈嘉、马铁军、周红艳、王立华、郭丛娜、李丹丹、刘梅、龙平、赵磊、马娟、熊洪营、王爱泽、吕盛蕾、陈炳基、李文娟、吴宏恩、林莉、袁倩、王山、濮丽佳、韩春笋、彭修文、章剑弘、李金龙、张萍、梁毅、林秀如、张淼、田多、金蕾蕾。

本文件为中国银行业协会制定,其著作权为中国银行业协会所有。

地 址:北京市西城区金融街20号交通银行大厦B座

电 话:010-66553368 010-66291132

邮 编:100033

邮 箱:cba.china@china-cba.net

传 真:010-66553356

引 言

在通过远程银行向银行客户提供优质服务的过程中，人工智能设施发挥着越来越重要的作用。而在目前的情况下，对人工智能客服的评价指标在不同的机构各有侧重，难以全面对人工智能客服在银行业务数字化转型中的作用进行完整和全面的评估。

本文件归纳了人工智能客服当前提供服务的实际情况，给出了一组在当前技术发展和业务现状下的人工智能评价指标。这些指标相对均衡地考虑了技术和业务的方面，给出的评价方法则是目前业界通用的评价方法。

本文件给出的对人工智能客服的评价方式，是从银行业金融机构内部视角出发的。实施本文件规定的评价，将对推进“线上化”业务实现数字化转型发展，推动人工智能客服能力持续提升，为客户提供智能、便捷、安全、优质的服务新体验，具有重要意义。

远程银行人工智能客服评价指标规范

1 范围

本文件规定了银行业远程银行与客服中心人工智能客服评价的指标范围及内容、指标体系与指标计算。

本文件适用于远程银行与客服中心人工智能客服能力和服务效果的评价与管理。

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 21664—2008 工作抽样方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能客服 AI customer service

以自动语音识别、自然语言理解和大数据运算等多种技术为基础，通过即时通讯工具，以语音、文字、图片等富媒体形式，为客户提供业务答疑、服务提醒、信息修改、消费交易等各类服务的设施。

术语条目注 1：人工智能客服目前主要包括文本服务机器人、语音导航机器人和智能外呼机器人。

3.2

文本交互AI AI on textual interaction

主要以文字方式响应服务请求的人工智能客服（3.1）。

术语条目注 1：文本交互 AI 支持文字交互、语音识别文字交互和图像识别文字交互等交互形式。

术语条目注 2：在银行业远程银行服务部门，往往称其为文本服务机器人。

3.3

语音交互AI AI on oral interaction

以语音方式响应服务请求的人工智能客服（3.1）。

术语条目注 1：在银行业远程银行服务部门，往往将应用于电话渠道，为客户提供菜单导航及语音问答交互服务的语音交互 AI 称为语音导航机器人；将通过电话或其它支持语音方式的交互，根据预设的业务场景流程自动发起语音外呼，收集业务结果或实现外呼业务目标的语音交互 AI 称为智能外呼机器人。

3.4

语音识别 speech recognition

由一个功能单元将语音信号转换为语音内容的表示形式

术语条目注 1：被识别的内容能够表示为一个适当的词汇或音素序列。

[来源：ISO/TR 24291:2021, 3.18, 有修改——未将“automatic speech recognition”作为许用术语]

3.5

银行客户 bank's customer

在银行经营的全生命周期中，已经或可能与银行发生业务关系的当事人。

术语条目注 1：当事人可能是自然人，也可能是法律认可的组织。

术语条目注 2：在本文件中，采用“客户”一词描述与银行联系的自然人或组织的代表。

[来源：GB/T 31186.1—2014，3.1，有修改——增加术语条目注2]

3.6

客户意图 customer's intent

银行客户（3.5）在与人工智能客服（3.1）的交互（3.7）过程中，通过语音、文字、图片及其他内容符号表达的希望达到某种目标的打算。

3.7

交互 transaction

由银行客户（3.5）与人工智能客服（3.1）之间的信息交换构成的活动最小单位

示例：人机一问一答、银行客户点击菜单后系统回复银行客户。

术语条目注 1：由银行业远程银行服务部门的视角看，交互往往指一次信息的输入及输出，如银行客户与人工智能客服之间信息输入输出的操作、交流、互动；交互方式包括文字交互、语音交互、图片交互、按键交互等。

术语条目注 2：在 ISO 30300:2020，3.1.29 的原文是“smallest unit of an activity consisting of an exchange between two or more participants or systems”。

[来源：ISO 30300:2020，3.1.29，有修改——增加术语条目注2]

3.8

会话 session

由银行客户（3.5）与人工智能客服（3.1）之间可执行多个交互（3.7）且持续维护交互数据的一段时间

术语条目注 1：由银行业远程银行服务部门的视角看，会话指银行客户与人工智能客服之间不中断的请求响应过程。通常情况下，会话数也称为进线量。会话结束的标志为客户明确结束会话或人工智能客服达到预设结束条件。

术语条目注 2：在 ISO/IEC TR 10032:2003，2.63 的原文是“a period of time during which a client may have many interactions with a server and both the client and server maintain data about each other”。

[来源：ISO/IEC TR 10032:2003，2.63，有修改——增加术语条目注2]

4 评价指标的范围、方法及内容**4.1 评价指标的范围**

4.1.1 性能指标，用于衡量人工智能客服语音识别及语义理解能力，主要包括语音识别准确率、问题识别率、意图理解准确率等。

4.1.2 运营效果指标，用于衡量人工智能客服在客户体验提升、服务分流等方面的能力，主要包括交互准确率、问题解决率、智能分流率等。

4.2 评价指标的方法**4.2.1 抽样法**

根据交互样本标注数据推算人工智能客服评价指标的方法。按照GB/T 21664-2008规定，假设事项发生率为50%、绝对误差为1%时，所需的必要样本量为1万。完成样本抽取后，与标注人员对本单位业务流程判定的结果进行对比，一致即为正确。

4.2.2 报表法

根据系统运营数据计算人工智能客服评价指标的方法。典型的系统运营数据包括进线量、交互数、客户评价数、转人工进线量。

4.3 评价指标的内容

4.3.1 语音识别准确率

4.3.1.1 指标释义

标注有效交互中，语音识别正确的字数在人工智能客服转写总字数中的占比。该指标主要适用于语音交互AI。

4.3.1.2 计算口径

4.3.1.2.1 适用语种

主要为中华人民共和国国家通用语言文字（普通话和规范汉字），包含阿拉伯数字及常用英文。在测算时，阿拉伯数字、单独英文字母（如：e缴费、ETC），一个数字或字母按一个文字计算；英文单词（如：card），一个单词按一个文字计算；标点符号不计算在内。

注：中华人民共和国国家通用语言文字必须符合《中华人民共和国国家通用语言文字法》的要求。

4.3.1.2.2 计算方法

语音识别准确率采用抽样法计算。

4.3.1.2.3 标注有效交互

标注有效交互指抽样标注交互中剔除无效语音后的交互。无效语音由标注人员在测听过程中参考判定规则进行判断，详见表1。

表1 无效语音判定规则

无效语音类型	释义
背景音	客户不是与语音交互 AI 对话，或旁边背景有人声，或在系统不支持打断的情况下，客户与机器人在同一时间说话的声音。
模糊音	客户发声比较模糊，导致标注人员无法准确标注。
杂音	周围环境存在杂音，用户说话声音过小，导致音频不清晰。
静音	无声音，标注人员听不到任何声音。
噪音	纯噪音，无人说话。如：嘟嘟声、电话接入前的“滴”声、公交车报站声、电视声音等。

4.3.1.3 计算公式

语音识别准确率按照如下公式计算：

语音识别准确率=（转写总数-替换字数-删除字数-插入字数）/转写总数×100%..... (1)

式中：

转写总数——语音交互 AI 转写出的文字的总字数；

替换字数——语音交互 AI 转写结果中的错误字数；

删除字数——系统未转写出的内容；

插入字数——系统额外转写的内容。

示例 1：如客户说“账户余额查询”，系统转写为“账户余额差询”，将“查”替换为“差”，则计为一个替换字数。

示例 2：如客户说“我想查一下账户还有多少钱”，系统转写为“我想查一下账户还有多少”，未转写出“钱”，则计为一个删除字数。

示例 3：如客户说“我想查一下账户还有多少钱”，系统转写为“我想查一下账户还有了多少钱”，额外转写“了”，则计为一个插入字数。

4.3.2 问题识别率

4.3.2.1 指标释义

人工智能客服已识别提问数在客户提问总数的占比。该指标主要适用于文本交互 AI。

4.3.2.2 计算口径

问题识别率采用报表法计算。

4.3.2.3 计算公式

问题识别率按照如下公式计算：

$$\text{问题识别率} = (\text{客户提问总数} - \text{未识别提问数}) / \text{客户提问总数} \times 100\% \dots\dots\dots (2)$$

式中：

客户提问总数——银行客户向文本交互 AI 提问的总数；

未识别提问数——包括文本交互 AI 未能匹配客户问题且未推送相似引导问的提问数之和。其中，相似引导问是指经系统计算，并与客户问题具有一定关联性的引导问题。

4.3.3 意图理解准确率

4.3.3.1 指标释义

人工智能客服对客户意图理解准确的交互数在语音识别正确交互数中的占比。该指标主要适用于语音交互 AI。

4.3.3.2 计算口径

意图理解准确率采用抽样法计算。

4.3.3.3 计算公式

意图理解准确率按照如下公式计算：

$$\text{意图理解准确率} = (1 - \text{标注意图理解错误交互数} / \text{标注语音识别正确交互数}) \times 100\% \dots\dots\dots (3)$$

式中：

标注意图理解错误交互数——语音识别正确的前提下，语音交互 AI 意图识别结果与人工标注意图结果不一致的交互条数；

标注语音识别正确交互数——标注有效交互中，所有文字语音识别正确的交互条数。

4.3.4 交互准确率

4.3.4.1 指标释义

人工智能客服根据客户意图，结合业务场景流程设计准确回应的情况。该指标主要适用于语音交互AI。

4.3.4.2 计算口径

交互准确率采用抽样法计算。

4.3.4.3 计算公式

交互准确率按照如下公式计算：

$$\text{交互准确率} = (1 - \text{标注交互错误数} / \text{标注有效交互数}) \times 100\% \dots\dots\dots (4)$$

式中：

标注交互错误数——标注有效交互中，语音交互AI实际交互结果与人工标注交互结果不一致的数量，包括因语音识别错误导致的交互错误，以及语音交互AI意图理解错误而导致的交互错误；

标注有效交互数——按照本文件 4.3.1.2.3 确定的有效交互。

4.3.5 问题解决率

4.3.5.1 指标释义

衡量人工智能客服解决客户问题的效果，以客户评价结果作为判断准则，即客户对人工智能客服“是否有效解决客户问题”的评价，该指标主要适用于文本交互AI。

4.3.5.2 计算口径

问题解决率采用报表法计算。

4.3.5.3 计算公式

问题解决率按照如下公式计算：

$$\text{问题解决率} = (1 - \text{客户评价未解决数} / \text{能评价客户提问总数}) \times 100\% \dots\dots\dots (5)$$

式中：

客户评价未解决数——客户与文本交互AI交互完成后，选择未解决或没帮助等选项的总数；

能评价客户提问总数——开通了评价功能渠道的客户提问数总和。

4.3.6 智能分流率

4.3.6.1 指标释义

在远程银行服务客户的过程中，由人工智能客服处理且未转人工的业务量在总进线量中的占比。

4.3.6.2 计算口径

智能分流率采用报表法计算。

4.3.6.3 计算公式

智能分流率按照如下公式计算：

$$\text{智能分流率} = (1 - \text{所有服务渠道转人工会话数} / \text{所有服务渠道进线会话总量}) \times 100\% \dots\dots\dots (6)$$

式中：

所有服务渠道转人工会话数——远程银行为客户提供服务的所有渠道（不含呼出）进入人工服务队伍的进线量；

所有服务渠道进线会话总量——远程银行为客户提供服务的所有渠道（不含呼出）的进线量总和，包含电话、移动终端或互联网页面等渠道。

5 综合评价指标权重

对人工智能客服提供服务的总体情况，在按照本文件第4章获取了相关的数据后，按表2给出的权重进行计算。

表 2 人工智能客服综合评价权重

指标类型	指标名称	AI 类型	权重 (%)
性能指标	语音识别准确率	语音导航机器人	5
		智能外呼机器人	5
	问题识别率	文本服务机器人	15
	意图理解准确率	语音导航机器人	10
智能外呼机器人		10	
运营效果指标	交互准确率	语音导航机器人	10
		智能外呼机器人	10
	问题解决率 ^a	文本服务机器人	15
	智能分流率	文本服务机器人及语音导航机器人	20
如何通过指标来改进服务，超出本文件的范畴。			
^a 此类指标主要站在客户视角来看待人工智能客服的绩效，其他指标主要站在银行视角来看待人工智能客服的绩效。			

参 考 文 献

- [1] GB/T 31186.1—2014 银行客户基本信息描述规范 第1部分：描述模型
 - [2] T/CBA 204-2019 远程银行客户服务与经营规范
 - [3] ISO/IEC TR 10032:2003 Information technology — Reference Model of Data Management
 - [4] ISO/TR 24291:2021 Health informatics—Applications of machine learning technologies in imaging and other medical applications
 - [5] ISO 30300:2020 Information and documentation — Records management — Core concepts and vocabulary
-