

T/CAICI

中国通信企业协会团体标准

T/CAICI XXXX—XXXX

5G 消息业务增强能力规范-AI 智能体能力 要求

5G Messaging Services Enhancement Capability Specification – Technical
requirements for AI Agent

(征求意见稿)

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

中国通信企业协会 发布

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
3.1 5G 消息 5G Messaging	1
3.2 聊天机器人 Chatbot	1
3.3 AI 智能体	1
3.4 5G 消息智能体	1
4 缩略语	1
5 5G 消息智能体技术要求总述	2
6 感知认知能力	2
6.1 对话交互	2
6.1.1 单轮对话	3
6.1.2 多轮对话	3
6.2 多模态感知	3
6.2.1 文本感知识别	3
6.2.2 位置感知识别	3
6.2.3 图片感知识别	3
6.2.4 语音感知识别	4
6.3 语义理解认知	4
6.3.1 文本纠错能力	4
6.3.2 意图理解能力	4
6.3.3 情感认知能力	4
6.4 推理能力	4
6.4.1 指代消解能力	4
6.4.2 知识推理能力	4
6.5 内容生成能力	4
6.5.1 文本生成能力	4
6.5.2 图片生成能力	5
6.5.3 位置检索生成能力	5
6.5.4 语音生成能力	5
6.5.5 AI 水印标识能力	5
6.5.6 富媒体卡片生成能力	5
7 规划能力	5
7.1 任务规划	5
7.1.1 目标拆解	5
7.1.2 规划策略	5
7.2 任务调度	5

7.2.1 优先级设置	5
7.2.2 调度机制	5
7.2.3 组织协调	6
7.3 任务优化	6
7.3.1 痕迹记录	6
7.3.2 结果评估	6
8 记忆能力	6
8.1 短期记忆	6
8.2 长期记忆	6
8.3 记忆检索	6
9 执行能力	6
10 底层模型要求	6
11 内生安全能力	7
12 5G 消息智能体终端入口要求	7
12.1 展示方式支持	7
12.2 交互功能支持	7
12.3 适配优化支持	7
12.4 系统权限配合	7

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。

本文件由中国通信企业协会团体标准管理委员会提出并归口。

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。

本文件起草单位：中国信息通信研究院、中国电信集团有限公司、中国移动通信集团有限公司、中国联合网络通信集团有限公司、中国通信企业协会增值服务专业委员会、中移互联网有限公司、联通在线信息科技有限公司、抖音视界有限公司、数海信息技术有限公司、深圳市壹通道科技有限公司、中邮世纪(北京)通信技术有限公司

本文件主要起草人：

本文件为中国通信企业协会首次发布。

5G 消息业务增强能力规范-AI 智能体能力要求

1 范围

本文件适用于行业客户或服务供应商建设的5G消息CSP平台及平台上承载的Chatbot应用,也同时适用于基础通信运营商提供的基于5G消息的智能体服务平台系统。本文件主要用于基于5G消息的AI智能体能力的技术要求和功能设计、开发、测试,也可用于对5G消息智能体情况进行监督、管理和评估。

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。

YD/T 3961-2021 5G消息 终端技术要求

3 术语和定义

下列术语和定义适用于本文件。

3.1 5G 消息 5G Messaging

5G消息是能与短彩信等基础电信业务协同开展的业务,5G消息业务包括个人消息、行业消息和增强通话涉及的消息等业务形式。

[来源: YD/T 3961-2021, 3.1.1]

3.2 聊天机器人 Chatbot

5G 消息中行业消息应用的呈现形式,以消息对话的方式,向用户提供行业消息服务功能。

[来源: YD/T 3961-2021, 3.1.2]

3.3 AI 智能体

AI智能体(又称AI Agent)是指人工智能代理(Artificial Intelligence Agent)是一种能够感知环境进行自主理解,进行决策和执行动作的智能体。AI Agent具备通过独立思考、调用工具逐步完成给定目标的能力。

3.4 5G 消息智能体

5G消息智能体是指在5G消息窗口中,基于5G消息的展现形式、遵循RCS UP规范,使用AI智能体技术强化Chatbot的智能化应用服务能力,提升用户交互体验和服务感知。

4 缩略语

下列缩略语适用于本文件。

GUD

全局用户标识

Global User Identifier

5 5G 消息智能体技术要求总述

5G消息智能体技术要求从感知认知、规划能力、记忆能力、执行能力、底层模型、内生安全六个方面进行规定；同时，针对手机终端5G消息智能体入口也进行相关技术要求提出。5G消息智能体的整体目标是在5G消息会话窗口中实现高效、智能、安全的交互体验，具体功能包括但不限于：

- a) 多模态输入输出支持：支持文本、图片、音频、视频、位置等多种模态的输入和输出。
- b) 智能对话管理：支持单轮和多轮对话，具备上下文管理能力，能够理解用户意图并进行连贯的对话。
- c) 多模态感知与理解：支持文本、图片、音频、位置等多种模态的感知识别与理解。
- d) 内容生成与展示：支持文本、图片音频、视频、位置、富媒体卡片等内容生成与展示，确保端到端的流程完整实现。
- e) 任务规划与执行：具备任务拆解、规划、调度、优化等能力，支持复杂任务的自动化处理。
- f) 记忆与检索：支持短期和长期记忆，能够快速检索历史信息，提升个性化服务能力。
- g) 安全与隐私保护：支持敏感内容检测，确保用户隐私和数据安全。

5G消息智能体能力不改变现有基础运营商的接口接入方式，可通过以下三种方式进行能力唤起：

- a) 通过“建议回复”的上行指令进行唤起：5G消息CSP平台接收到用户“建议回复”中反馈的唤起指令，将用户的会话服务切换到5G消息智能体接口服务中。
- b) 通过用户上行文本或语音输入进行唤起：5G消息CSP平台接收到用户上传的文本内容或者语音ASR转换后的文本进行意图识别，如果识别出唤起指令，则将用户的会话服务切换到智能体接口服务中。
- c) 通过独立Chatbot做为5G消息智能体入口：用户可以通过独立的Chatbot应用直接唤起智能体服务，该独立的Chatbot支持在智能交互中唤起其他Chatbot应用进行服务。

5G消息CSP平台可具备关键词回复（含正则等模糊匹配机制）、传统NLP回复、智能体回复（含通用大模型交互）三种人机交互能力。如果具备多于一种交互能力，应针对不同的Chatbot应用具备能力优先级配置和调度能力。

针对AI大模型和5G消息智能体应用的交互，应采用非流式模式进行用户回复反馈。同时，若接入的智能体服务非流式回复响应超过1秒，应提供友好的用户等待提示词（如：思考中、检索到xx条资料内容整理中、工具调用中等），缓解用户焦虑等待，提升用户体验。

在5G消息会话中，5G消息智能体生成的富媒体内容（图片、音频、视频）需自动先进行运营商素材送审，审核通过后再进行下行消息体的组装和下发。禁止未经审核的AIGC内容对公众用户进行推送。

5G消息CSP平台可接入私有化部署或第三方智能体能力，为保障用户手机号码等敏感数据，需在与智能体服务后台交互时对用户手机号码进行脱敏，采用全局用户标识（Global User Identifier，简称GUID）方式进行用户识别。同时，智能体服务应用应独立进行用户会话的管理和策略处理。

6 感知认知能力

6.1 对话交互

6.1.1 单轮对话

5G消息智能体应支持与用户的智能对话交互，具体要求包括：

- a) 可通过5G消息对话框，支持用户文本、图片、音频、视频、位置五种模态输入，并充分利用5G消息的富媒体能力，提供更丰富的交互体验；
- b) 可围绕用户对话内容进行自动化回应，回应内容应支持文本、图片、音频、视频、位置、富媒体卡片等多媒体形式，确保在5G消息平台上的展示效果；
- c) 支持的文本输出长度不超过2000字；
- d) 响应时间超过1秒时，建议提供实时反馈提示（如“正在处理中，请稍候”）。

6.1.2 多轮对话

除单轮对话外，5G消息智能体还应支持与用户的多轮智能对话交互，具体要求额外包括：

- a) 具备上下文会话管理能力，包括保存用户历史输入、关键意图和已完成任务的状态，能够对用户提问中的省略语或指代（如“上一条”“那个”）进行解析，确保对话逻辑的连贯性；
- b) 可快速响应用户在多轮对话中的意图切换；用户意图模糊或难以识别时，可通过追问或建议问题的方式主动引导用户明确需求；
- c) 原则上应支持无限轮对话，为优化性能和用户体验建议默认支持至少10轮连续对话，并在接近限制时提示用户。

6.2 多模态感知

6.2.1 文本感知识别

5G消息智能体应支持对用户输入文本的感知识别能力，具体要求包括：

- a) 应支持多种格式和结构的文本输入解析，包括但不限于短句、长文本、缩写、口语化表达；
- b) 应具备处理多种语言、方言及地区性差异的能力，确保在不同语言环境下都能进行有效的文本理解和处理；
- c) 应具备根据不同场景任务自动切换文本识别模式的能力，适应多场景化的需求，如证件识别、票据识别、表格识别、自定义文档识别等；
- d) 能够识别用户输入中的关键字、意图及情感色彩；
- e) 应支持错误处理与反馈，文字无法识别或识别错误时，提供清晰的错误提示；

6.2.2 位置感知识别

5G消息智能体应支持对用户输入位置的感知识别能力，具体要求包括：

- a) 应支持不同形式的位置信息输入，包括用户手动输入的地址、位置分享等；
- b) 能够准确解析用户提供的地理位置信息，包括但不限于GPS坐标转换、地址文本、地名等；
- c) 应支持错误处理与反馈，位置无法识别或识别错误时，提供清晰的错误提示。

6.2.3 图片感知识别

5G消息智能体应支持对用户输入图片的感知识别能力，具体要求包括：

- a) 应支持多种类型图片的识别和解析，包括但不限于：包含文字的图片、场景图片、物体图片、二维码/条形码等；
- b) 应支持多种格式图片的识别和解析，包括但不限于jpg、jpeg、png等；
- c) 应支持图片的内容分析，包括主体检测、颜色分析、场景分析等；
- d) 应支持错误处理与反馈，图片无法识别或识别错误时，提供清晰的错误提示。

6.2.4 语音感知识别

5G消息智能体应支持对用户输入语音的感知识别能力，具体要求包括：

- a) 应支持语音转文本（ASR）功能，能够准确识别清晰语音，并处理包含噪音或口音的语音输入；
- b) 应具备处理多语言的能力，能够支持多种语言和方言（例如，普通话、粤语、英语等），并根据用户的语言习惯进行语音识别；
- c) 应具备语音中的意图识别和情感分析能力，识别用户的需求、问题或命令，并根据情感状态调整响应方式；
- d) 应支持错误处理与反馈，语音无法识别或识别错误时，提供清晰的错误提示。

6.3 语义理解认知

6.3.1 文本纠错能力

5G消息智能体应支持对用户输入错误文本及语音转文本错误的自动修改与校正，应能准确识别用户输入的拼写错误、语法错误以及语音转文本过程中产生的识别错误。同时，5G消息智能体应提供多样化的纠错建议，包括但不限于直接更正文本、提供正确的词汇或短语替换，以及上下文相关的修正建议。

6.3.2 意图理解能力

5G消息智能体应能够准确理解并解析用户的要求和指令，可对用户提出的包含多个步骤或条件的复杂指令进行准确解析，并理解其内在的逻辑和顺序。5G消息智能体应能够基于对话的上下文信息，准确理解用户当前意图，并能够适应不同用户群体和场景的意图表达习惯，包括方言、俚语、行业术语等。

6.3.3 情感认知能力

5G消息智能体应能够准确判断并理解用户的情感极性，能够准确区分用户的正面情感（如高兴、满意）、负面情感（如生气、失望）以及中性情感。

6.4 推理能力

6.4.1 指代消解能力

5G消息智能体应能够准确理解和解析对话中的指代关系，确保对话的连贯性和准确性。

6.4.2 知识推理能力

5G消息智能体应具备增强知识检索、数学计算以及逻辑推理能力。

6.5 内容生成能力

6.5.1 文本生成能力

5G消息智能体应具备高效、智能的文本生成能力，能够基于用户输入的上下文生成自然流畅、符合逻辑的文本响应，并支持多语种识别和生成。5G消息智能体应具备关键信息提取和摘要能力。同时，5G消息智能体可根据用户的情绪或对话情境调整应答风格。

6.5.2 图片生成能力

5G消息智能体应具备生成和处理图片的能力，可根据用户输入的文本描述生成图片，并理解用户持续改进意图，进行图片的智能优化。5G消息智能体生成的图片尺寸和规格应遵循《5G消息业务显示规范》标准要求。

6.5.3 位置检索生成能力

5G消息智能体应具备位置检索与生成能力，支持依据用户意图进行实时地理信息检索、路线规划等，可根据用户历史位置数据与偏好生成个性化服务推荐。

6.5.4 语音生成能力

5G消息智能体应具备语音生成能力，将文本信息转换为自然流畅的语音输出，可根据文本内容调整语音的语速、语调、音量，以适应不同的场景和用户需求。同时，可支持多种语言的语音生成，包括中文、英文等。

6.5.5 AI 水印标识能力

5G消息智能体生成的AI内容应参照网信办《人工智能生成合成内容标识办法(征求意见稿)》和国家标准《网络安全技术 人工智能生成合成内容标识方法》要求，具备AI内容的水印标识功能。

6.5.6 富媒体卡片生成能力

5G消息智能体应支持富媒体卡片的生成与展示，支持单卡片、轮播卡片生成，展示结构化的多媒体形态消息。同时，生成多媒体分辨率应按照5G消息展示规范要求，在不同设备上都可以适配完整渲染。

7 规划能力

7.1 任务规划

7.1.1 目标拆解

5G消息智能体应在理解用户意图后，对用户指定的目标及任务进行拆解，使拆解后的目标任务更易实现和执行。

7.1.2 规划策略

5G消息智能体应支持任务内或任务间的组织，支持并行、迭代等多种结构的任务规划。并对拆解后的任务进行合理性的评估，保障任务间合理协调运行。

7.2 任务调度

7.2.1 优先级设置

5G消息智能体应具备任务优先级设置功能，允许用户根据实际情况对任务优先级进行动态调整，并对不同优先级的任务自动进行合理的资源分配。

7.2.2 调度机制

5G消息智能体应支持多种任务调度机制，如先来先服务、短作业优先等，且支持按照预先设置的优先级进行任务调度。在面对异常状况时，5G消息智能体应具备重新规划任务调度的机制。

7.2.3 组织协调

5G消息智能体在执行任务时，应具备资源协调、任务分配、进度监控、应急处理等协调能力。

7.3 任务优化

7.3.1 痕迹记录

5G消息智能体应具备痕迹记录功能，可记录并存储用户与智能体之间的交互对话关键信息，并提供安全可靠的存储机制，保障用户数据隐私，防止数据泄露或被非法访问。

7.3.2 结果评估

5G消息智能体应具备任务规划结果的评估能力和反馈优化能力，可基于评估结果和用户反馈，进行自我反思能力，通过算法分析自身表现，识别改进空间。

8 记忆能力

5G消息智能体的应实现智能体与用户交互信息的高效的存储与检索，并确保系统的性能与用户体验的可靠性。

8.1 短期记忆

短期记忆用于跟踪用户当前任务或上下文至少5轮对话信息，确保对话和操作的连贯性，一般适用于即时响应的场景。

8.2 长期记忆

5G消息智能体应支持存储长期的、有持久价值的信息，存储知识库或信息档案。比如记住用户的常见问题和答案，通过用户交互生成用户画像，支撑后续进行用户的个性化服务。

8.3 记忆检索

5G消息智能体可根据上下文关键词检索、模糊匹配快速找到相关记忆的结果。

9 执行能力

5G消息智能体应支持与其他智能体、系统、环境本身等进行交互，支持流程理解、任务分配、异常处理和多重工具调用，保证流程的正常运行。

10 底层模型要求

5G消息智能体在底层模型能力方面应支持对主流大模型的适配，具体要求包括：

a) 5G消息智能体应提供标准的API接口，以便于与大模型进行高效、稳定的数据交互和指令传输。同时，建议支持私有化部署方式，以满足不同客户对数据安全和隐私保护的需求。

b) 5G消息智能体应能够适配多种类型的大模型，包括大语言模型、视觉模型、多模态模型等多种底层模型；

c) 底层大模型必须通过符合国家标准《人工智能 大模型 第1部分：通用要求》，确保其商用可行性和技术可靠性。

11 内生安全能力

5G消息智能体应支持对生成文本、图像内容的安全检测，具体要求包括：

- a) 支持涉黄、涉政、涉暴敏感类别检测，建议支持对广告、违禁、诈骗类别的检测；
- b) 文本检测应支持主流的字符编码格式；
- c) 应支持中文的检测，建议支持包含英文等多语种的检测；
- d) 图片检测应支持主流图片格式，包括jpeg、gif、png；
- e) 图片检测应支持对图片中文本的检测。

12 5G 消息智能体终端入口要求

为让用户更好地体验5G消息智能体，建议终端厂商在支持5G消息智能体入口方面，需从展示方式、交互功能以及适配优化等多维度进行配合，推动5G消息智能体的广泛应用。

12.1 展示方式支持

- a) 消息列表置顶：终端如支持官方5G消息智能体入口展示，建议将5G消息智能体助手固定置于消息列表顶部展示。
- b) 应用号主界面置顶：在应用号主界面中，建议将5G消息智能体助手置顶展示。
- c) 支持短信负一屏下拉展示：建议终端通过短信负一屏下拉的方式，对5G消息智能体以及AI工具集成广场进行展示，方便用户快速搜索5G消息智能体，找到自己喜欢的AI工具。

12.2 交互功能支持

悬浮球语音唤醒：建议终端结合语音识别算法，实现悬浮球语音唤醒5G消息智能体助手功能。

12.3 适配优化支持

5G消息智能体流式输出支持：建议终端进行5G消息智能体流式输出适配，即5G消息智能体在生成回复内容时，能够逐字、逐句或逐段实时显示在终端界面上，而不是让用户等待完整回复生成后才看到内容，让用户在等待过程中就能了解5G消息智能体的处理进度，增强交互的实时感和流畅感，减少用户等待焦虑。

12.4 系统权限配合

为配合5G消息智能体生活助手的各种场景使用，建议终端配合5G消息智能体获取必要的系统权限，如通讯录、位置信息等，终端可在获取全显示给予用户必要的权限申请提示，运营商需对相关数据加密传输存储，保障用户隐私与数据安全。
