

Original data

Dataset of user-agent conversations

坐席：您好请问有什么可以帮您？
(Agent:Hello, how can I help you?)

客户：你给我查137****7276的手机话费
(User: You check my cell phone bill for 137****7276)

坐席：好的是本机套餐吗先生？
(Agent : Okay is that the local package sir?)

客户：嗯(User : Uh-huh.)

坐席：请您稍等，您现在当前的话话费余额有三十四块五。
(Agent : Please wait a moment, you now have a balance of thirty-four dollars and five cents in your current phone bill.)

Dataset Complaint Order

“省：江西赣州[地址] 业务号码 *****3004,用户来电反馈，不显示5G信号，用户其它电信号码显示____，只有此号不显示，终端也是支持5G，要求核实回复，谢谢。”
("Province: Ganzhou, Jiangxi [address] business number *****3004: Users call feedback does not show 5G signal, the user other telecom numbers show, only this number does not show, the end of the Rui is also support for 5G, asked to verify the response, thank you.)

Data desensitization

Defining Desensitized Fields

Setting Desensitized Characters

Cell Phone No.: [Cell Phone No.]
Name: last name [first name]
Address: [Address]
Other numbers: ****
.....

Data Cleaning

Elimination of non-significant characters/spellings

Space/Underline/...

Manual checking

Expert verification

AI Expert Verification

Answer/Label Construction

1

Labeling of experts (Example: Complaint order classification)

Input:“省:江西赣州[地址] 业务号码*****3004,用户来电反馈，不显示5G信号，用户其它电信号码显示，只有此号不显示，终端也是支持5G，要求核实回复，谢谢(Province: Ganzhou, Jiangxi [address] business number *****3004,users call feedback, does not show 5G signal, the user other telecom numbers show, only this number does not show, the end of the Rui is also support for the 5G, asked to verify the response, thank you!)

Output: 投诉单→网络质量类→无信号→无信号
(Complaint List→Network Quality Category→No Signal→No Signal)

2

Screening data based on keywords provided by experts (Example : Tendency Analysis of network transfer

Input:

#坐席:您好请问有什么可以帮您？

#客户:你好我想问下电信的携号转网现在有什么优惠套餐吗？

#坐席:女士.那我帮您转入坐席给您处理这个问题可以吗？

#客户:可以。

#坐席:好的您稍等请不要挂机。

#客户:嗯。

#坐席:请不要挂机正在帮您转接中。”

Output: 1 (Tendency to switch to other networks)

#Agent:Hello, how can I help you?

#User:Hello, I would like to ask if there are any special packages available for the telecom's number portability.

#Agent:Madam, can I transfer you to operator to deal with this problem?

#Customer:Yes. # User :Yes.

#Agent: Okay, please wait a moment.

#User :Uh-huh.

#Attendant:Please do not hang up, we are transferring you.”)

3

GPT Generation: Generate dialogs based on scene descriptions (Example: Intelligent Order Filling)

Prompt1: “以运营商装维场景为例，生成一段客户和坐席人员仿真对话数据，对话中包含约定装维上门时间和询问用户地址消息，其中地址为随机生成的真实地址精确到小区”

Output1:

坐席人员：您好，我是[运营商]的坐席人员，很高兴为您服务。
用户：最近我的宽带信号一直不稳定，网速变慢，有时候还会断网....。

坐席人员：非常抱歉，我会尽量帮您解决问题。请问方便告诉我您所在的地址吗？

用户：我住在北京市海淀区西二旗，中关村软件园三期，15号楼。

坐席人员：好的.我会尽快为您安排上门检修。请问您什么时间比较方便呢？

用户：最好是周末的上午，因为平时我上班没有时间。

坐席人员：好的，那我们安排在本周六上午10点为您进行上门维修。

用户：那太好

坐席人员：再见！祝您有一个愉快的周末！

Prompt2:将上述文本中的地址和时间抽取出来

Output2:

地址：北京市海淀区西二旗，中关村软件园三期，15号楼。

时间：本周末上午10点左右。

Take the operator installation and maintenance scenario as an example, generate 1 segment of user and agent simulation dialog data, the dialog contains the agreed installation and maintenance door-to-door time and ask for the user's address message, in which the address is a randomly generated real address accurate to the cell'.

Agent : Hello, I am the seat staff of [operator], glad to serve you.

User: Recently, my broadband signal has been unstable, the network has become slower, and sometimes there is a disconnection

Agent: I am very sorry, I will try to help you solve the problem. Can you tell me your address?

User: I live in No.15 Building, Zhongguancun Software Park Phase III, Xi'erqi, Haidian District, Beijing.

Agent: Yes, I will arrange for you to come to my home as soon as possible. What time is more convenient for you?

User: It's best to do it in the morning on weekends, because I don't have time to go to work during weekdays.

Agent: Okay, we will arrange for your door-to-door repair at 10:00 am on Saturday.

User: That's great!

Agent : Goodbye! Wish you have a pleasant weekend!

Extract the address and time from the above text

Address: Building 15, Zhongguancun Software Park, Phase III, Xi'erqi, Haidian District, Beijing.

Time: about 10:00 am this weekend.

原始数据

坐席用户对话数据集

坐席：您好请问有什么可以帮您？

客户：你给我查137****7276的手机话费

坐席：好的是本机套餐吗先生？

客户：嗯

坐席：请您稍等，您现在当前的话费余额有三十四块五。

.....

客户：谢谢您了

...

投诉工单数据集

"省：江西赣州 _ _ _ _ _ 业务号码
*****3004：__用户反映 国际漫游当天25元但在24小时就会扣钱50多元，客户需要告知怎么扣款的，加急解决问题请尽快处理并回复，谢谢！联系电话199****3004：*****（请分公司按照用户要求，了解用户实际办理情况，与用户解释，48小时内尽快回复）请省公司确认是否涉及新羽行动（2G\3G退网）"

数据脱敏

定义脱敏字段

手机号：[手机号]
姓名：姓[姓名]
地址：[地址]
其他数字：***
.....

设置不同类型的脱敏字符

数据清洗

剔除无实际意义字符/错别字

空格/下划线/...

人工核对

业务专家核对

NLP技术人员核对

答案/标签构建

①

专家标注（投诉工单分类任务）

Input:“用户反映 国际漫游当天25元但在24小时就会扣钱50多元，客户需要告知怎么扣款的，加急解决问题 请尽快处理并回复，谢谢！联系电话[手机号码（请分公司按照用户要求，了解用户实际办理情况，与用户解释，48小时内尽快回复）请省公司确认是否涉及新羽行动（2G\3G退网）”

Output: 查询单→服务协调类→省自定→渠道服务

②

基于业务专家提供的关键词筛选对话（携号转网倾向性分析任务）

Input:“坐席：您好请问有什么可以帮您？#客户：哎你好我想问下电信的携号转网现在有什么优惠套餐吗？#坐席：是这样的女士那你我帮您转入我们的专家坐席给您处理这个问题可以吗？#客户：可以。#坐席：好的您稍等请不要挂机。#客户：嗯。#坐席：请不要挂机正在帮您转接中。”

Output: 1（具有携号转网倾向性）

③

GPT生成：基于场景描述ChatGPT生成类似场景对话（智能填单）

Prompt1: ‘以运营商装维场景为例，生成1段客户和坐席人员仿真对话数据，对话中包含约定装维上门时间和询问用户地址消息，其中地址为随机生成的真实地址精确到小区’

Prompt2: 将上述文本中的地址和时间抽取出来

Output:

坐席人员：您好，我是[运营商]的坐席人员，很高兴为您服务。

用户：最近我的宽带信号一直不稳定，网度变慢，有时候还会断网....。

坐席人员：非常抱歉，我会尽量帮您解决问题。请问方便告诉我您所在的地址吗，

用户：我住在北京市海淀区西二旗，中关村软件园三期，15号楼。

坐席人员：好的，我会尽快为您安排上门检修。请问您什么时间比较方便呢？

用户：最好是周末的上午，因为平时我上班没有时间。

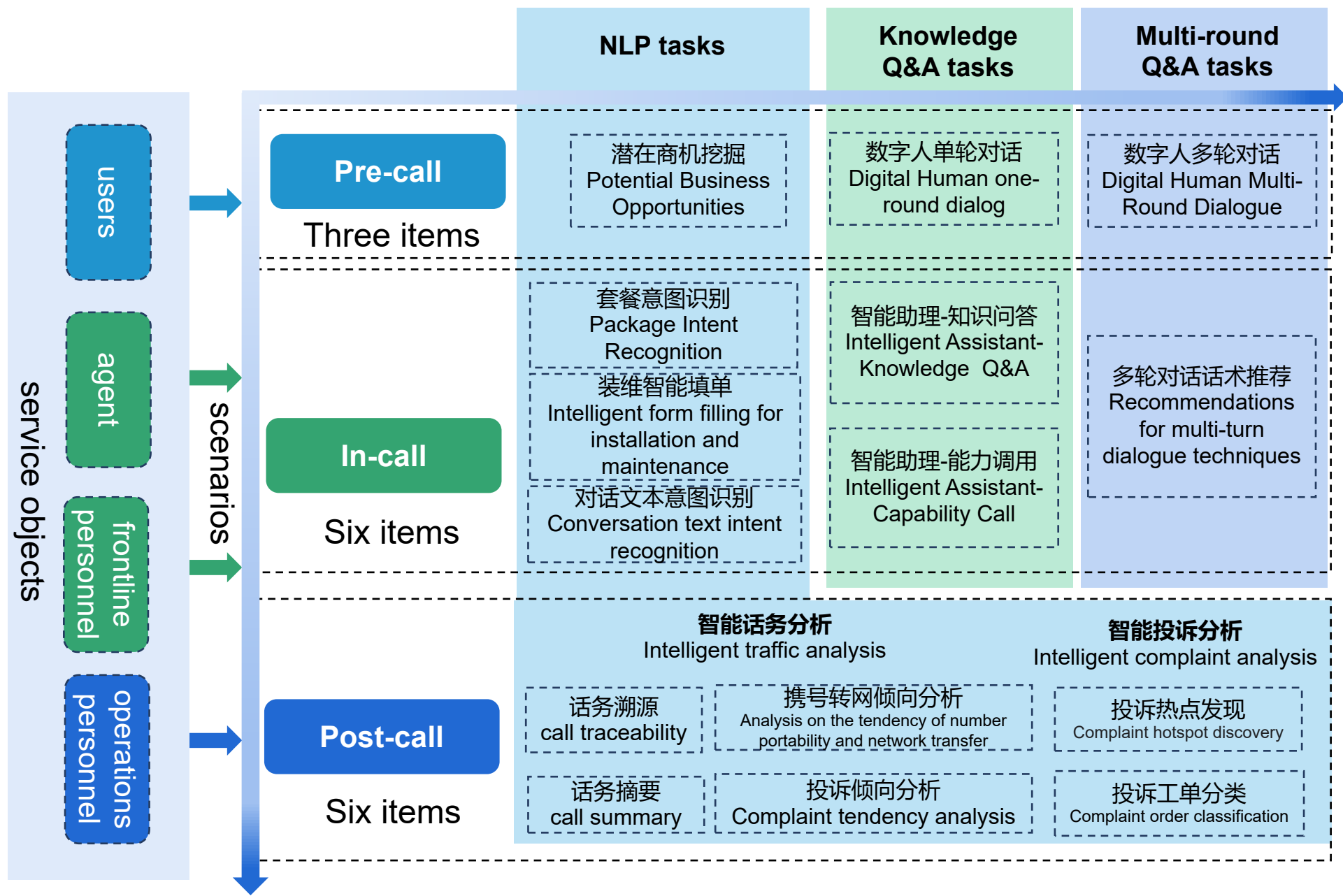
坐席人员：好的，那我们安排在本周六上午10点为您进行上门维修。

用户：那太好

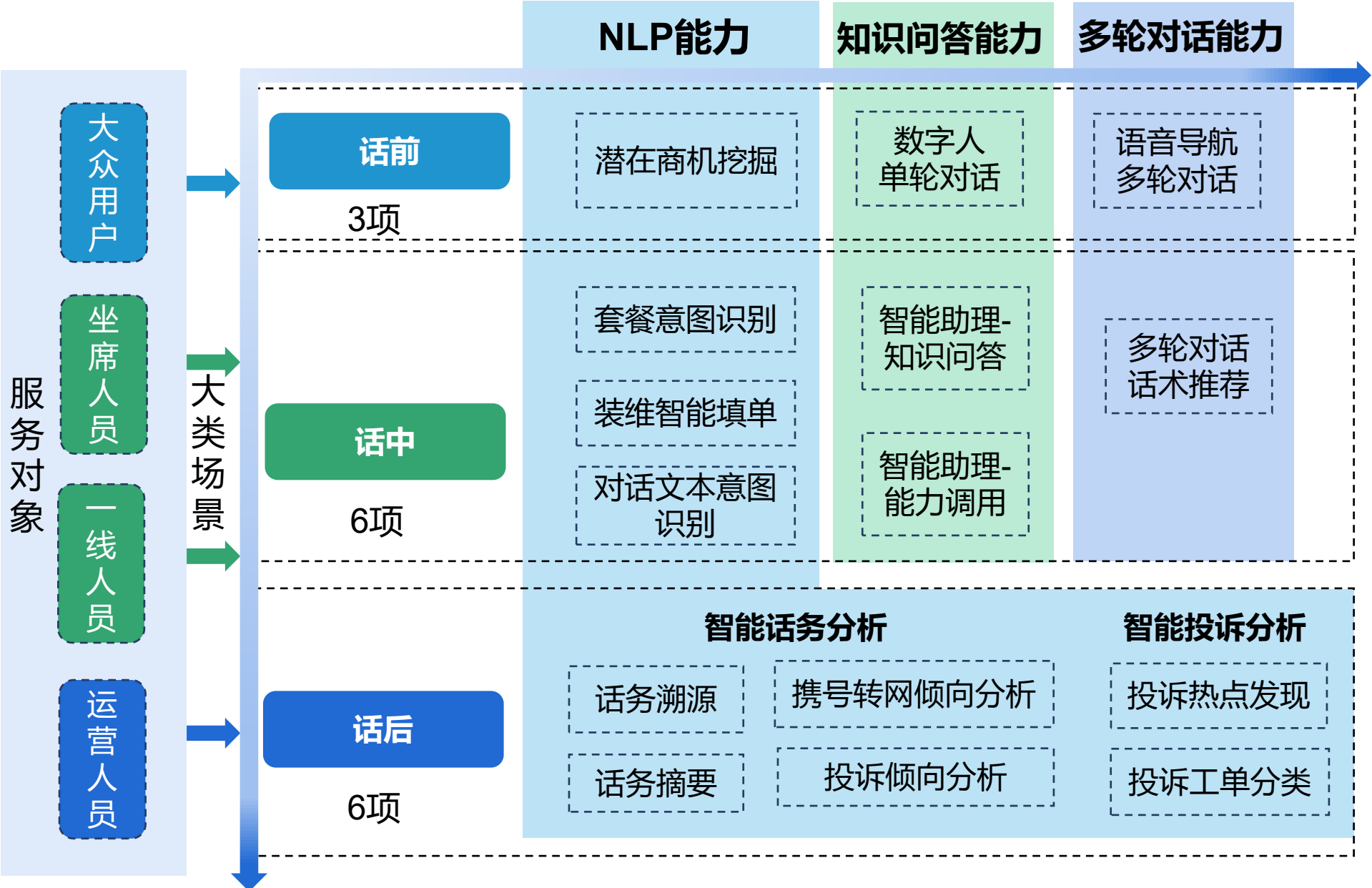
坐席人员：再见！祝您有一个愉快的周末！

地址：北京市海淀区西二旗，中关村软件园三期，15号楼。

时间：本周末上午10点左右。



客服场景评估体系

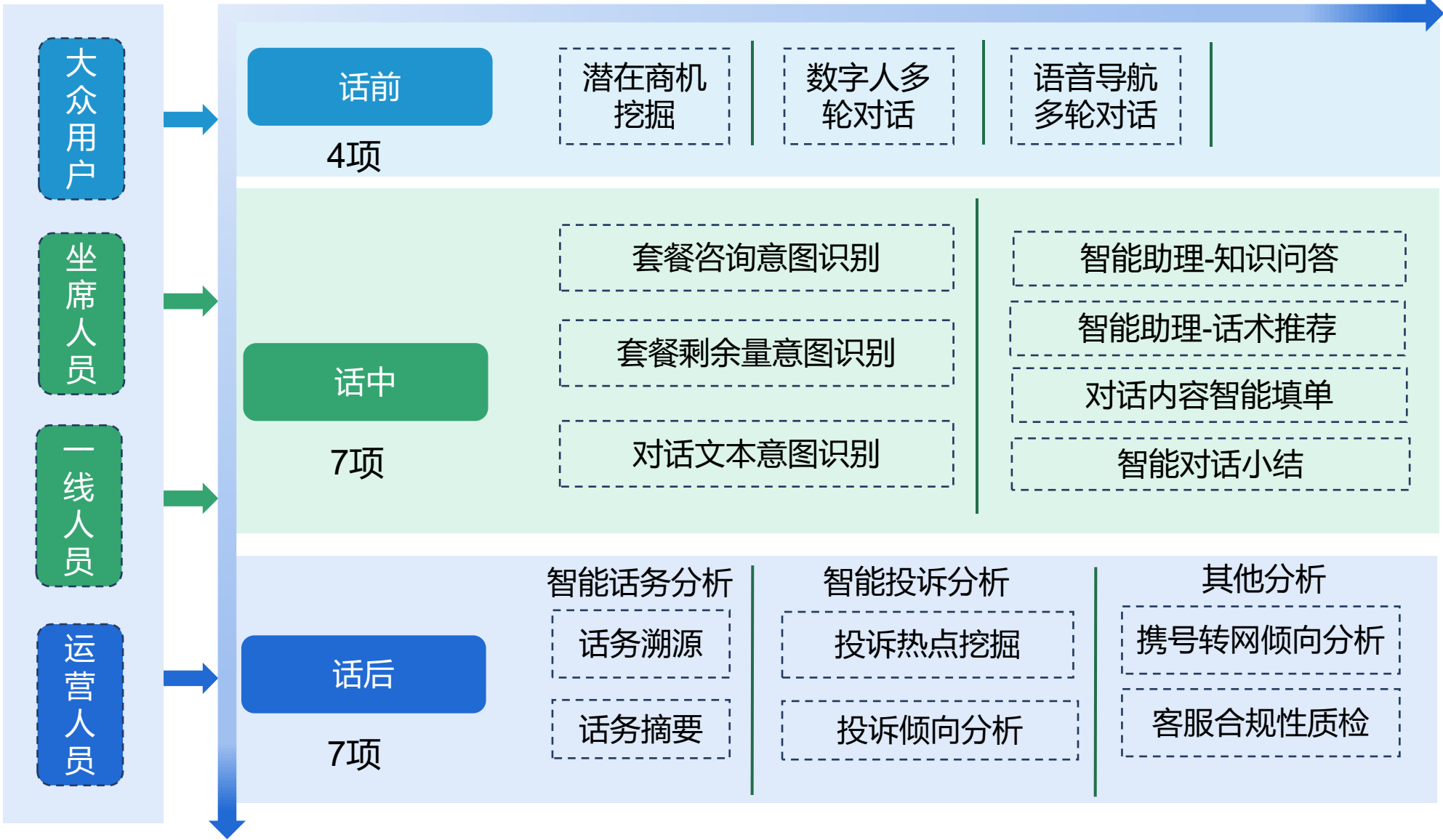


客服场景评估体系

服务对象

大类场景

小类场景

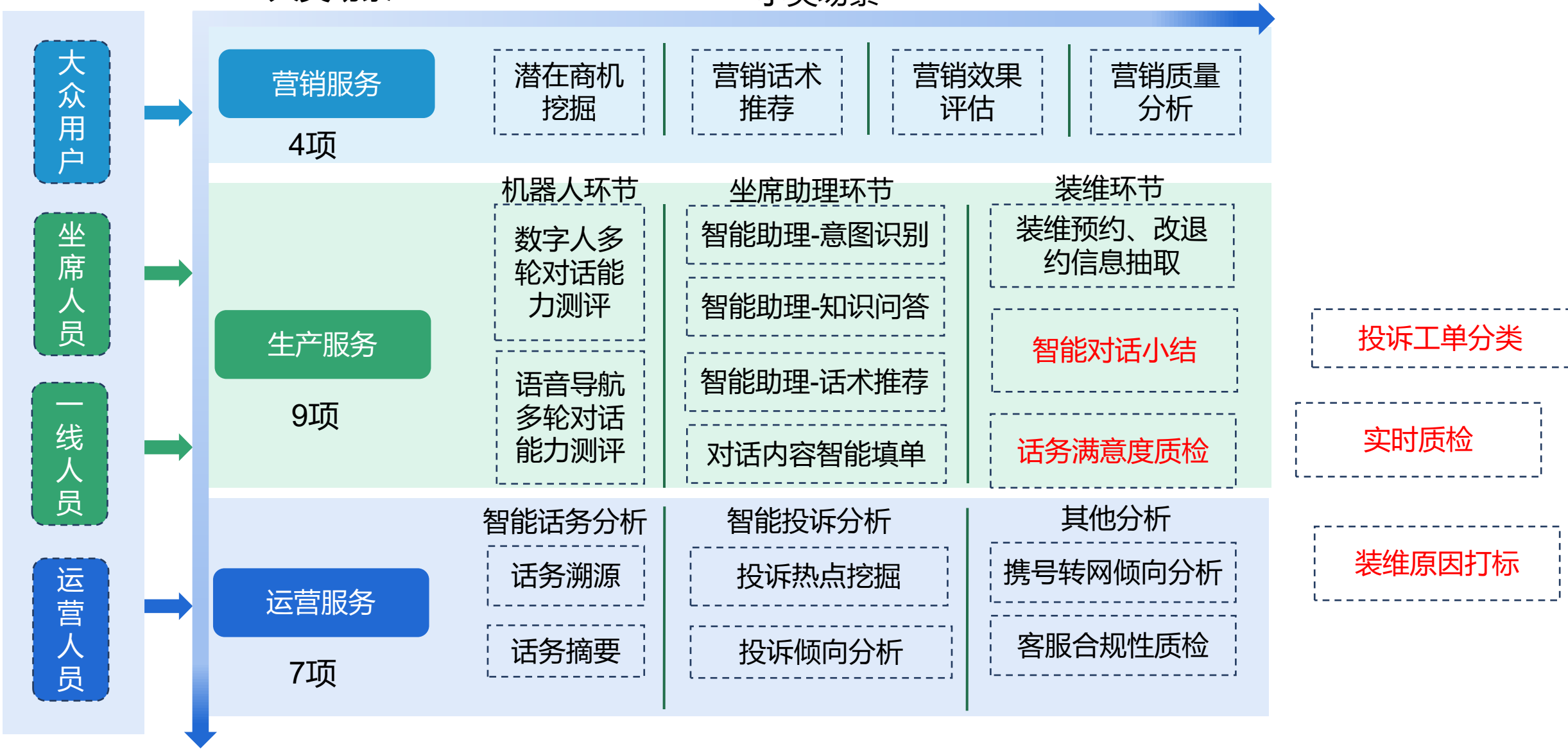


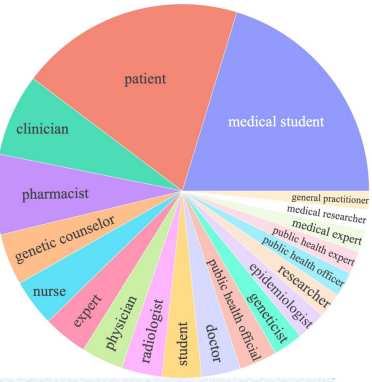
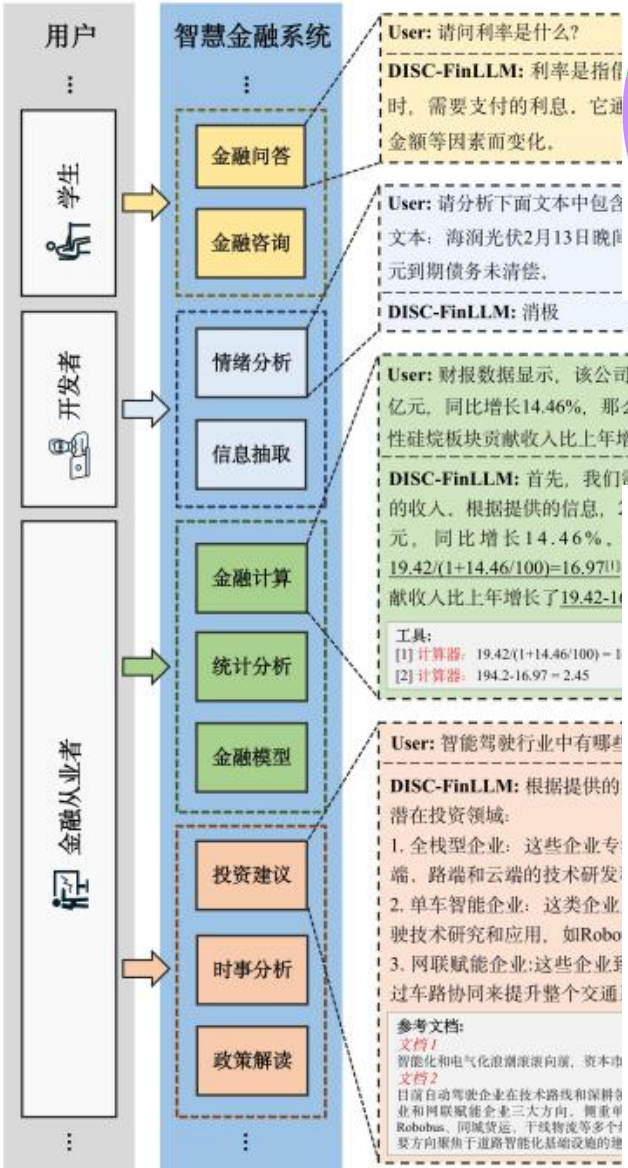
客服场景评估体系

服务对象

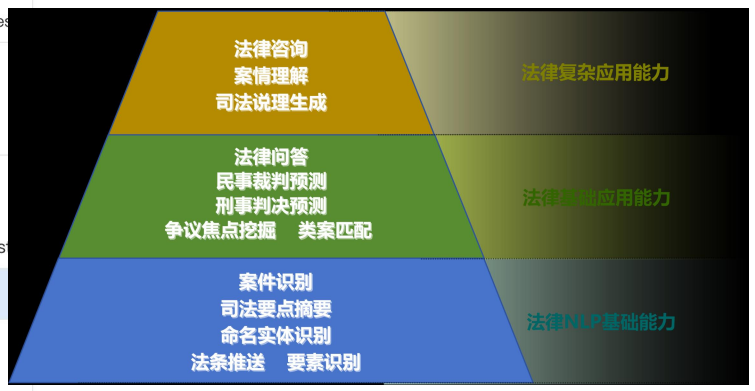
大类场景

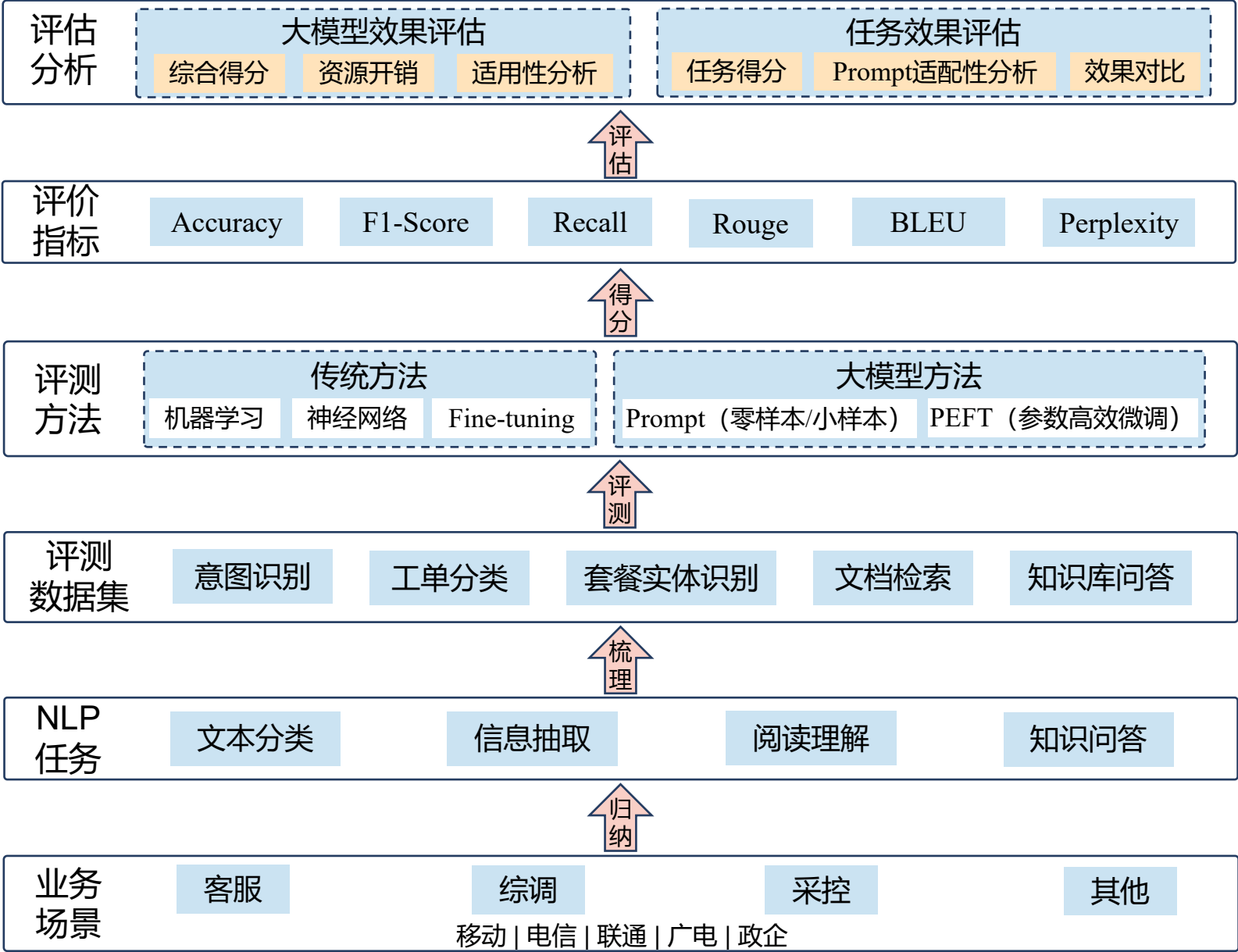
小类场景





PLAN		CODE		BUILD		TEST	
REQM	ProgM	Design	FrameWorks	SCM/VCS	PackagMgmt	UnitTest	IntegrationTest
RiskMgmt		General		DBMgmt		SystemTest	UAT
		Programming		DepMgmt		PerformanceTest	SecurityTest
RELEASE		DEPLOY		OPERATE		MONITOR	
ArtifactsMgmt	CI/CD	ConfigMgmt	Provisioning	Containerization	Virtualization	Data	Alert
Linux		Cloud		OperateSense		Visualization	
		IAC		AIOps		Analysis	





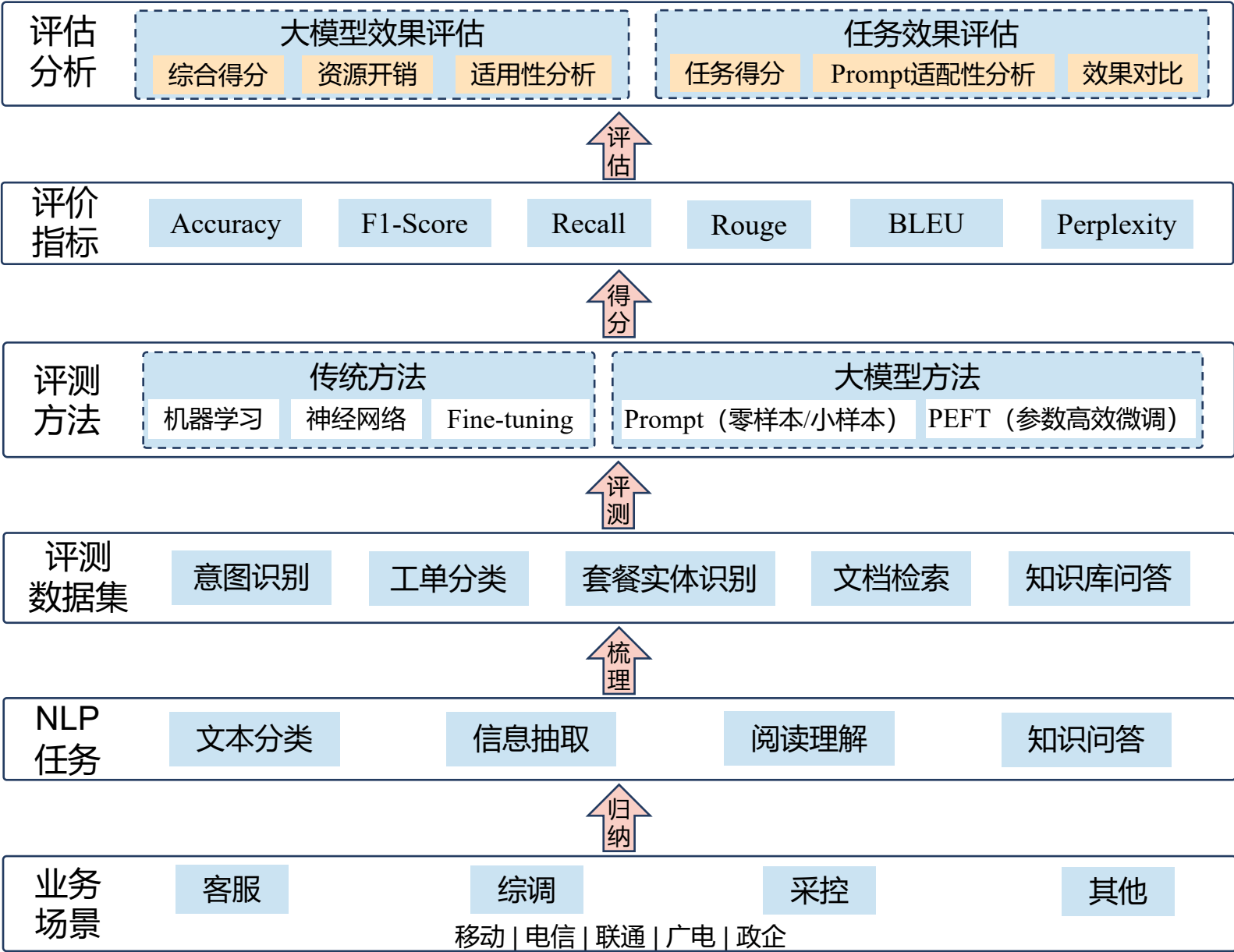
基于公司运营商行业 **3** 大类业务场景，对应 NLP 任务梳理行业评测数据集 **20+**，在 **30+** 开源大模型上进行了广泛的评测。

挑战:

- 大模型存在**事实性错误**、知识盲区、常识偏差、**生成内容不可靠**、内容安全性等问题
- 大模型发展迅速，评测体系滞后，并且在**行业应用上缺乏完整可靠的评估标准**

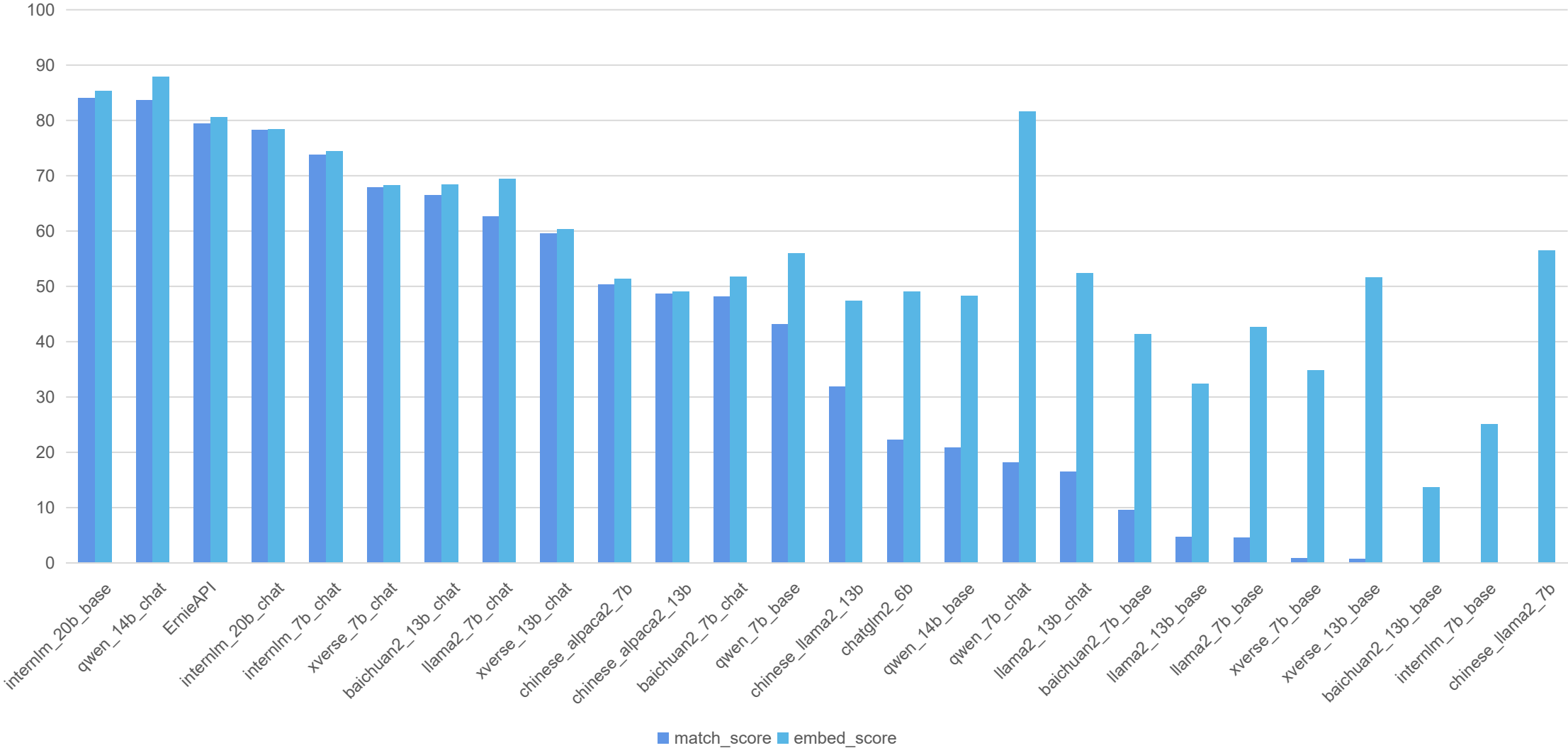
意义:

- 根据场景任务梳理行业数据，**建立标准的运营商评估数据集**
- 汇总评测方法，研究任务适配的Prompt工程，**建立统一的可控生成框架**
- 形成完整可靠且自动化的运营商大模型评测体系
- 促进行业大模型的研发，把握成本控制，促进大模型应用落地





客服场景-原始大模型



评估场景
4大类

评估能力
4大类

评估场景
20+

客服

- 文本分类能力
 - 套餐意图识别
 - 宽带问题分类
 - ...
- 信息抽取
 - 套餐实体识别
 - 套餐属性抽取
- 知识问答
 - ...
- 开放域问答
 - ...

采控

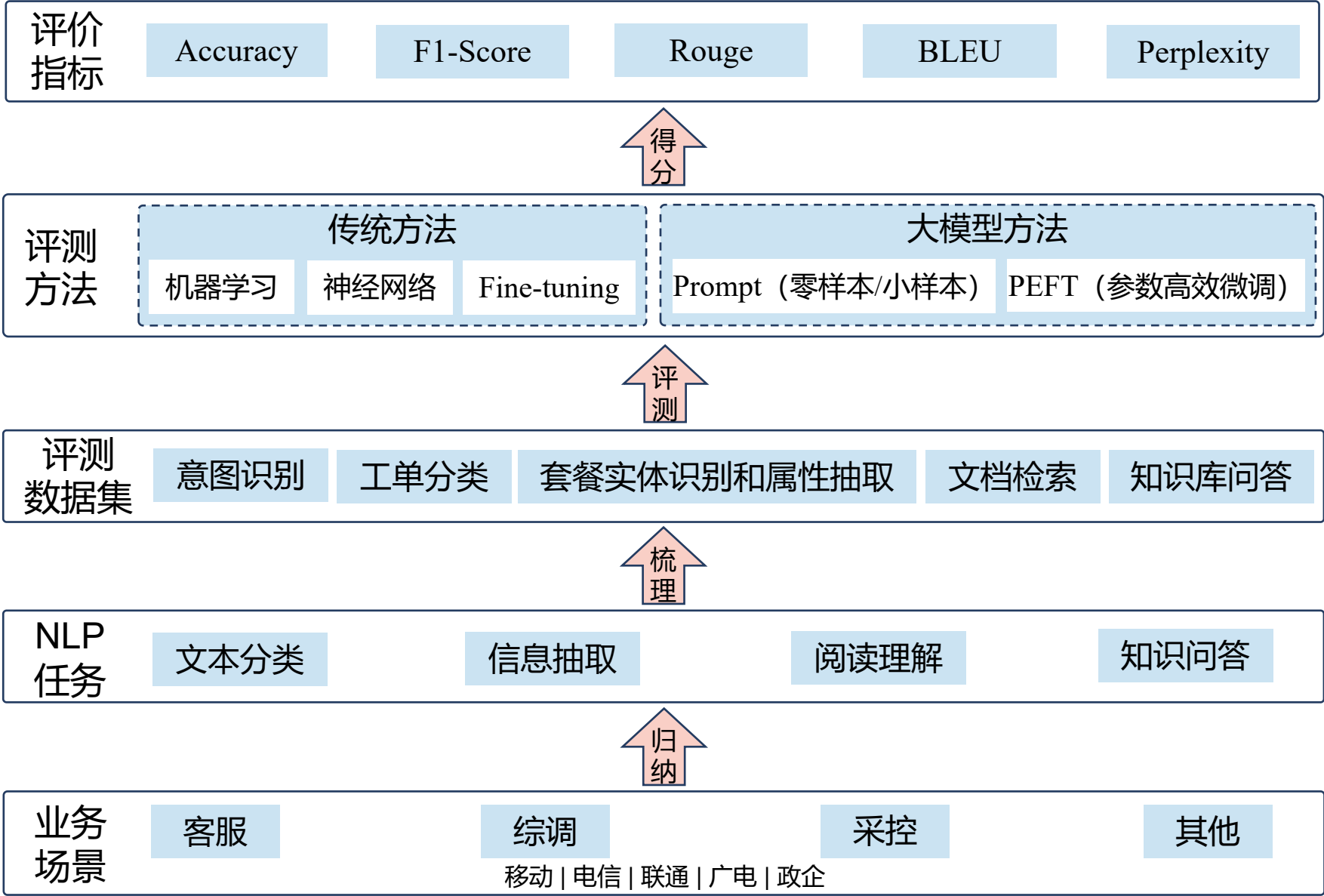
- 文本分类能力
 - 网络告警类型
 - 网络告警等级
 - ...
- 信息抽取
 - 故障事件抽取
 - 故障三元组抽取
- 知识问答
 - ...
- 开放域问答
 - ...

综调

- 文本分类能力
 - 工单环节预测
 - 装维工单分类
 - ...
- 信息抽取
 - 装维地址分级
 - ...
- 知识问答
 - ...
- 开放域问答
 - ...

其他

- 文本分类能力
 - 政务群众意图分类
 - ...
- 信息抽取
 - 天气预警信息抽取
 - 农业知识抽取
- 知识问答
 - ...
- 开放域问答
 - ...



挑战:

- 大模型存在事实性错误、知识盲区、常识偏差、生成内容不可靠、内容安全性等问题
- 大模型发展迅速，评测体系滞后，并且在行业应用上缺乏完整可靠的评估标准

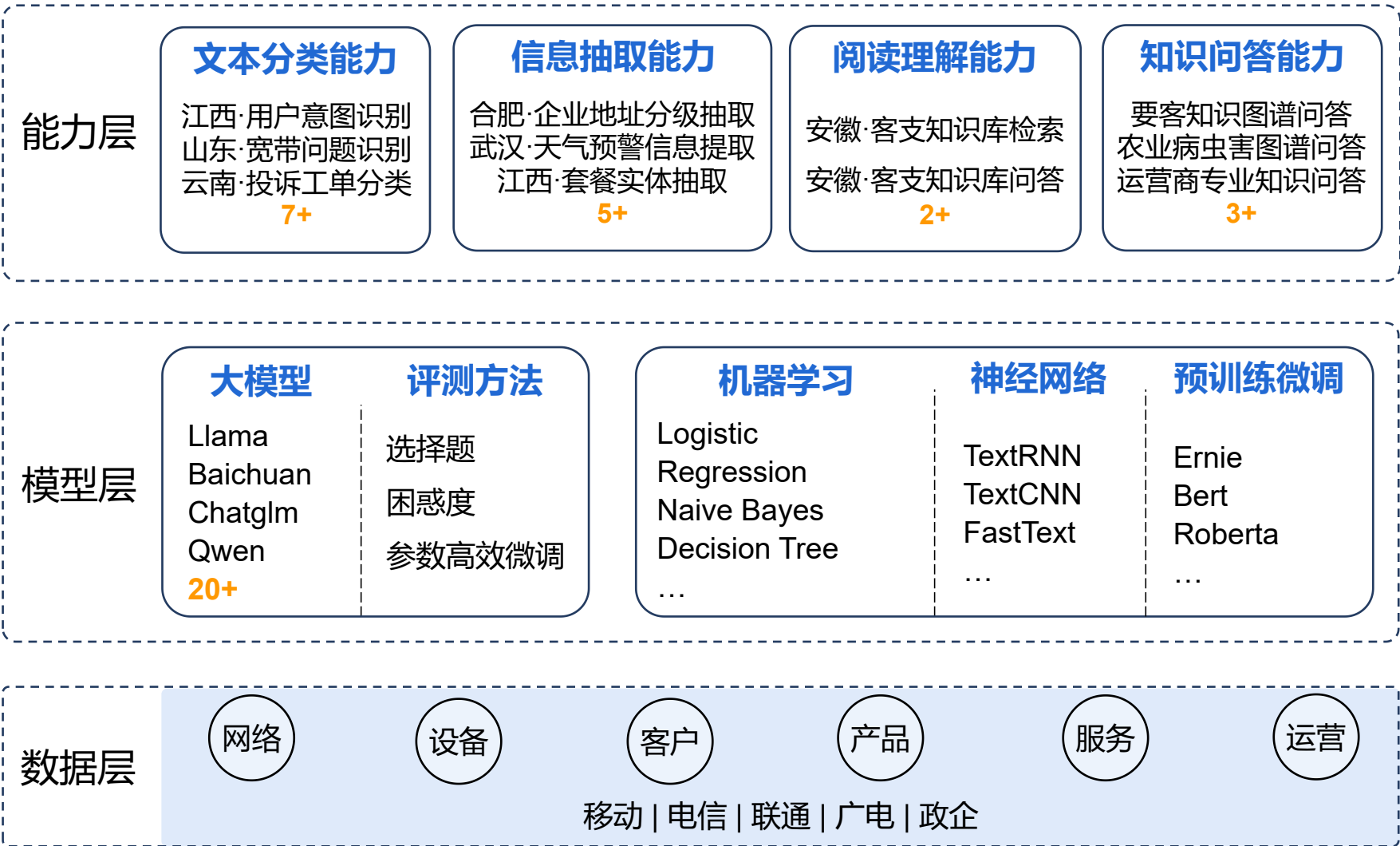
意义:

- 根据场景任务梳理行业数据，建立标准的运营商评估数据集
- 汇总评测方法，研究任务适配的Prompt工程，建立统一的可控生成框架
- 形成完整可靠且自动化的运营商大模型评测体系
- 促进行业大模型的研发，把握成本控制，促进大模型应用落地

挑战

大模型存在事实性错误、知识盲区、常识偏差、生成内容不可靠、内容安全性等问题

大模型发展迅速，评测体系滞后，并且在应用上缺乏行业大模型评估标准



意义

行业应用

模型选择

参数分析

可控生成

成本评估

安全可靠

标准体系

研究改进

