

云智未来<sup>9</sup><sup>th</sup>

第九届中国系统架构师大会  
SYSTEM ARCHITECT CONFERENCE CHINA 2017

# 可定制开发的语音交互技术

— DUI的全链路定制探索之路

张顺@思必驰

SACC  
2017

北京·新云南皇冠假日酒店

IT168.com

ChinaUnix

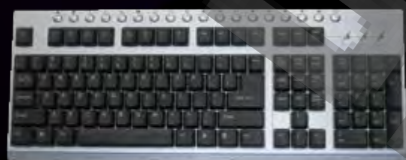
ITPUB

# 主要内容

- 01) 语音交互的探索经历
- 02) DUI的可定制的语音交互
- 03) 一些问题与思考

# PART 1

## 语音交互探索经历



键盘 1873



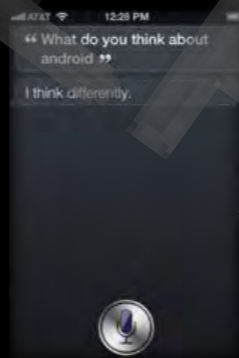
鼠标 1968



Xerox Alto 1973  
最早的图形化系统



iPhone 2007  
多点触摸应用于手机



Siri 2011  
语音助手应用于手机



Amazon Echo 2014  
语音音箱应用

AI SPEECH | UI



命令行



PC时代



图形交互



移动时代



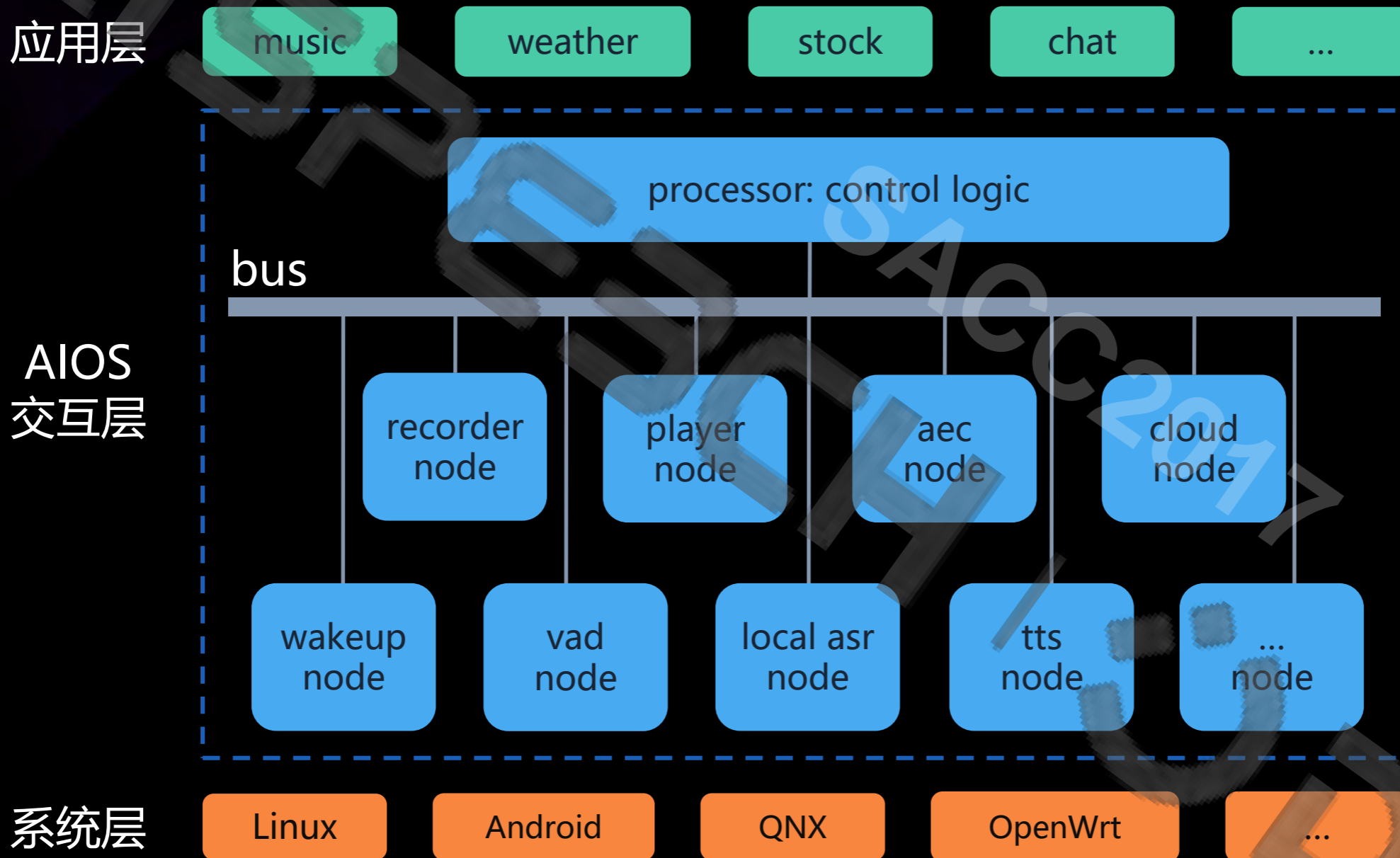
语音交互



iOT时代

- a. **产品方面**：一个app一个语音入口, 各种语音助手交互方式各异, 非刚需
- b. **工程方面**：研发难度大, 基于SDK/API自己实现需要了解语音对话相关知识, 自己开发
- c. **语音技术**：识别性能逐年突破





\* 由AIOS实现各场景下的完整的交互逻辑

\* 与底层操作系统和应用层解耦

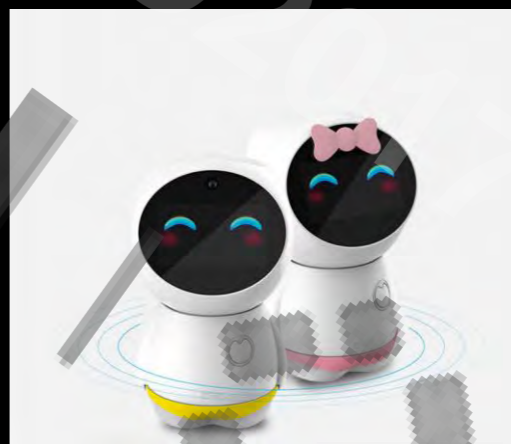
- a. **产品方面**：刚需的车载, 家居, 机器人等
- b. **工程方面**：完整的交互系统, 集成简单, 定制能力不足, 出现一些语音系统方案商
- c. **语音技术**：识别性能逐步突破, 一个垂域使用一个模型, 严重依赖语音厂商

特殊词语识别，新词等识别  
同一个说法，不同解析要求  
对话交互逻辑个性化定制  
不同产品共用一个模型资源，融合多家需求，更新难

我要听<凉凉>

打电话前是否要确认?

我要听郭德纲





The screenshot displays the DUI platform's skill development interface. At the top, there are navigation tabs: 总览, 产品, 技能, and 资源. The current task is '天气'. The interface includes a '意图' (Intent) list on the left and a '语义槽' (Semantic Slot) table at the bottom.

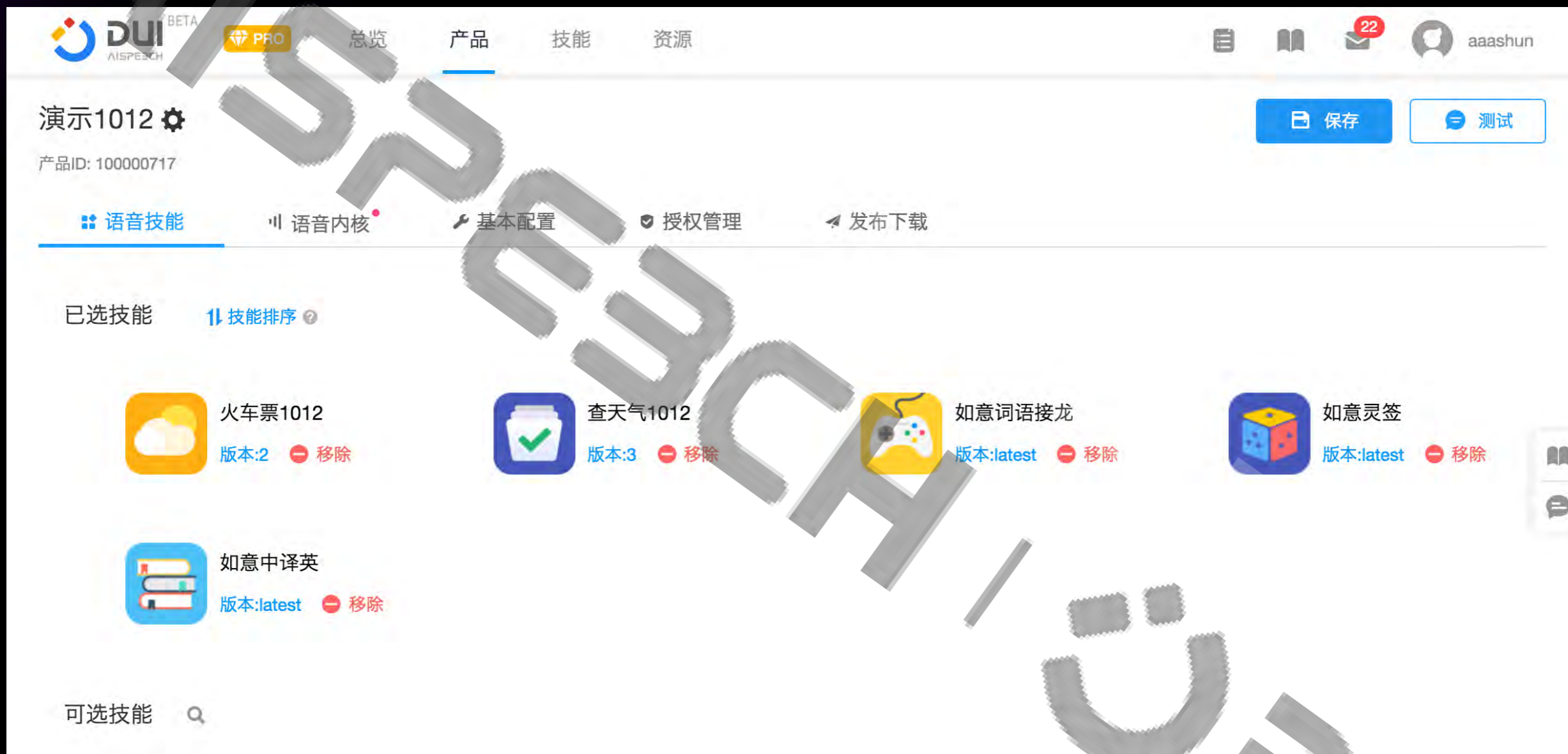
**意图 (Intents):**

- #日期-[和|跟|与]#日期#[的]天气
- #日期#[和|跟|与]#日期-[的]天气
- #日期#[是](什么|啥|什么样的|啥样的|啥样|什么样|哪种)天气
- #城市#[是](什么|啥|什么样的|啥样的|啥样|什么样|哪种)天气

**语义槽 (Semantic Slots) Table:**

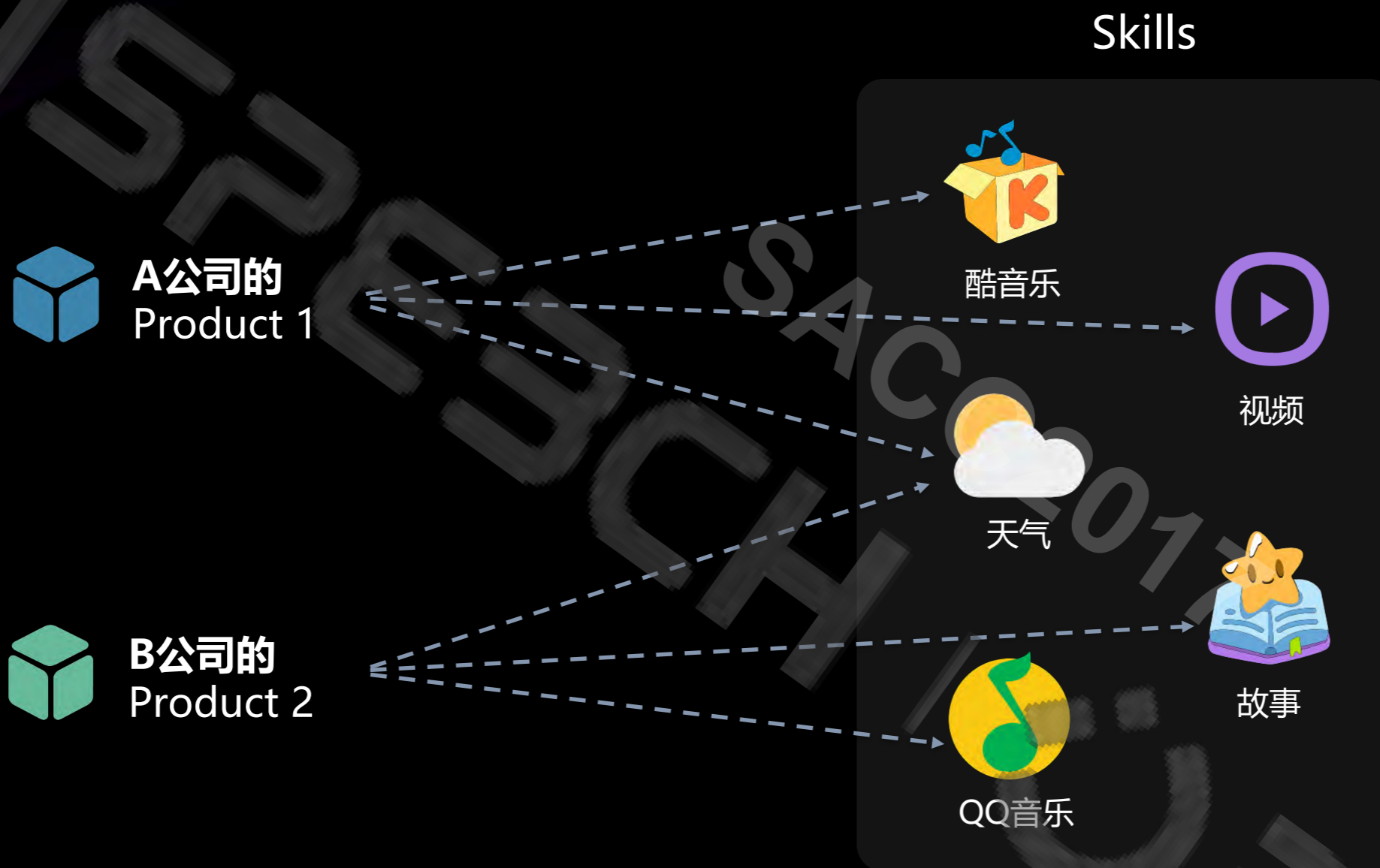
顺序	必需	名称	词库	提问
-	<input type="checkbox"/>	城市	@sys.国内城市	
-	<input type="checkbox"/>	日期	@sys.日期	
-	<input type="checkbox"/>	查询对象	@sys.天气查询对象	
-	<input type="checkbox"/>	区域	@sys.国内市辖区	

**Skill开发(类似App开发)**  
对Skill开发者更**简单**, 可自助式定制开发



### 产品接入

更简单快速地接入新产品, 勾选使用语音Skill

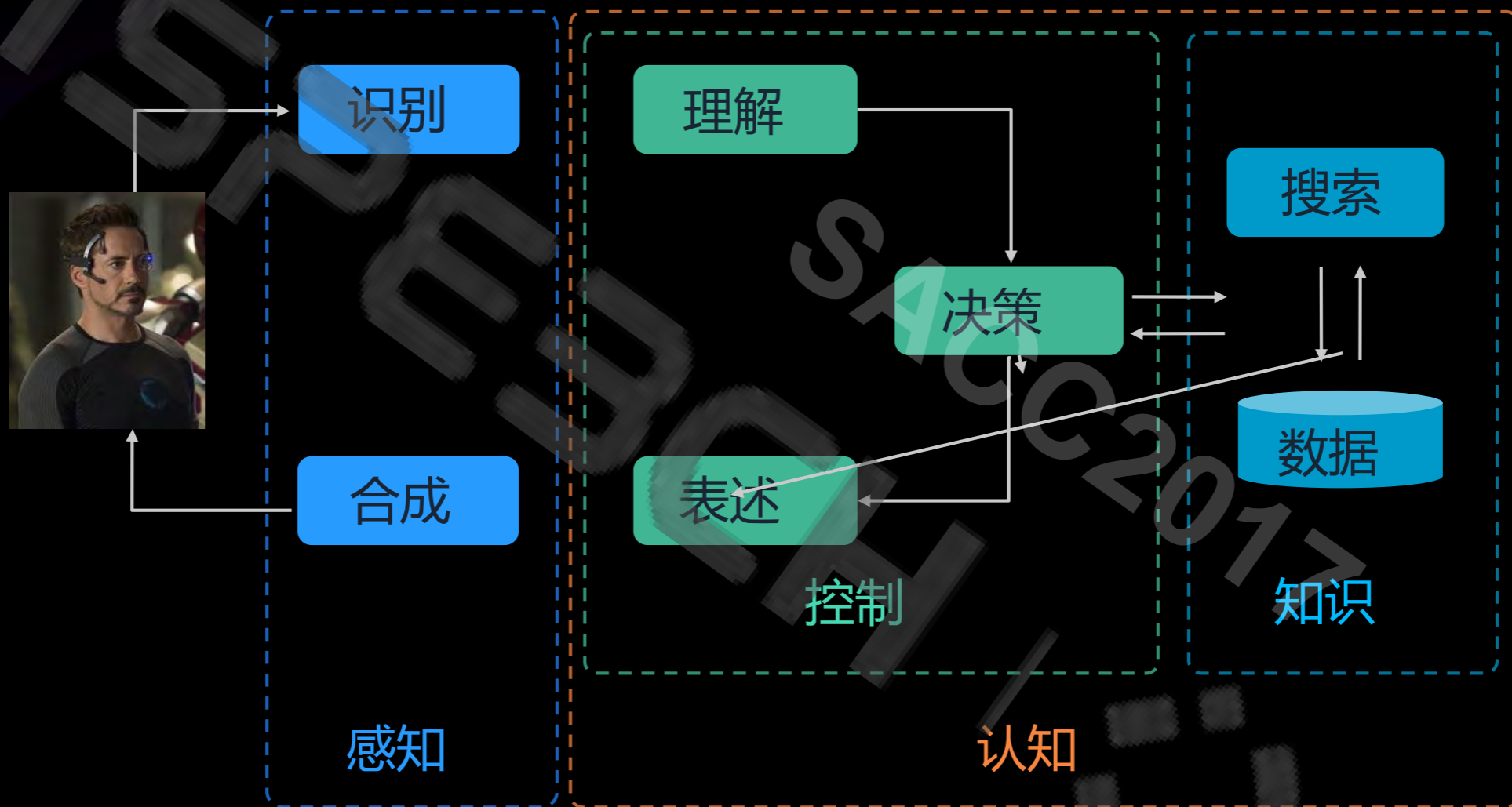


Skill即是一个完整的可交互的单位, 类似App  
Skill开发过程涉及到识别,语义,对话等的定制开发

PART 2

# DUI的可定制的语音交互

# DUI的可定制的语音交互



语音唤醒

语音识别  
语音合成

语义理解  
对话管理  
知识库



检索型

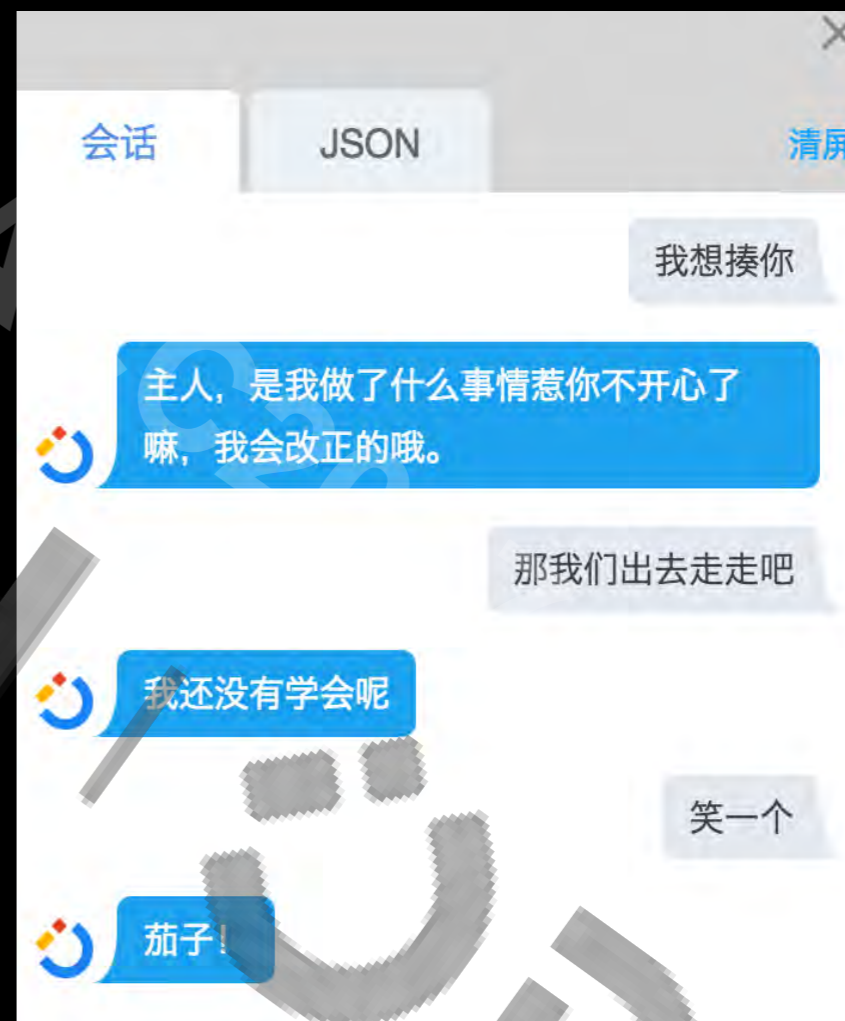


知识型



任务型

- 非结构化内容，不好形成知识库
- 一般是单轮一问一答
- 典型应用是聊天



Q : 你怎么这么聪明  
A : 我可是天生滴

Q : 不洗手能吃饭吗  
A : 小朋友要洗手



北京天气不错啊

语义分析

检索

排序

我也这么觉得

### 基于信息检索的Skill



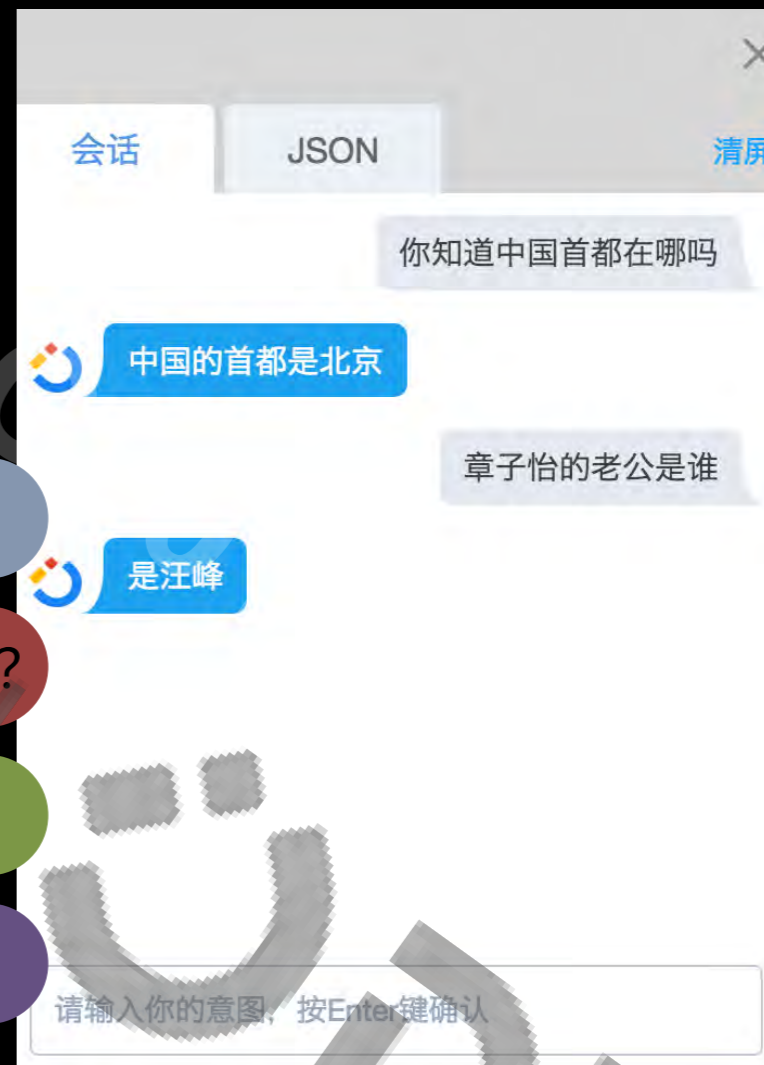
- 特定域
- 知识可以结构化
- 一般是单轮一问一答
- 用户期待较高

章子怡的老公的年龄？

美国总统是谁？

三生三世十里桃花是谁主演的

思必驰有哪几个部门





### 基于知识图谱的Skill

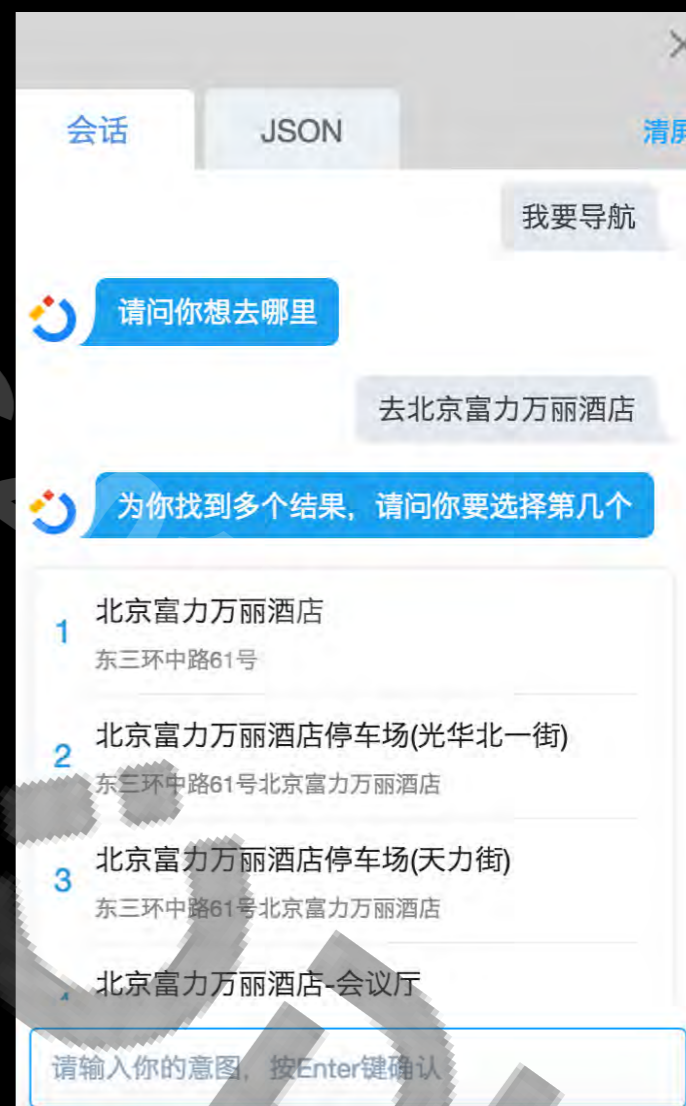
- 特定域
- 带着目的来的, 比如导航, 打电话, 听歌
- 多轮对话, 但是轮数越少越好
- 需要跟踪对话状态以及决策策略
- 严肃场景, 用户期待非常高, 不容出错

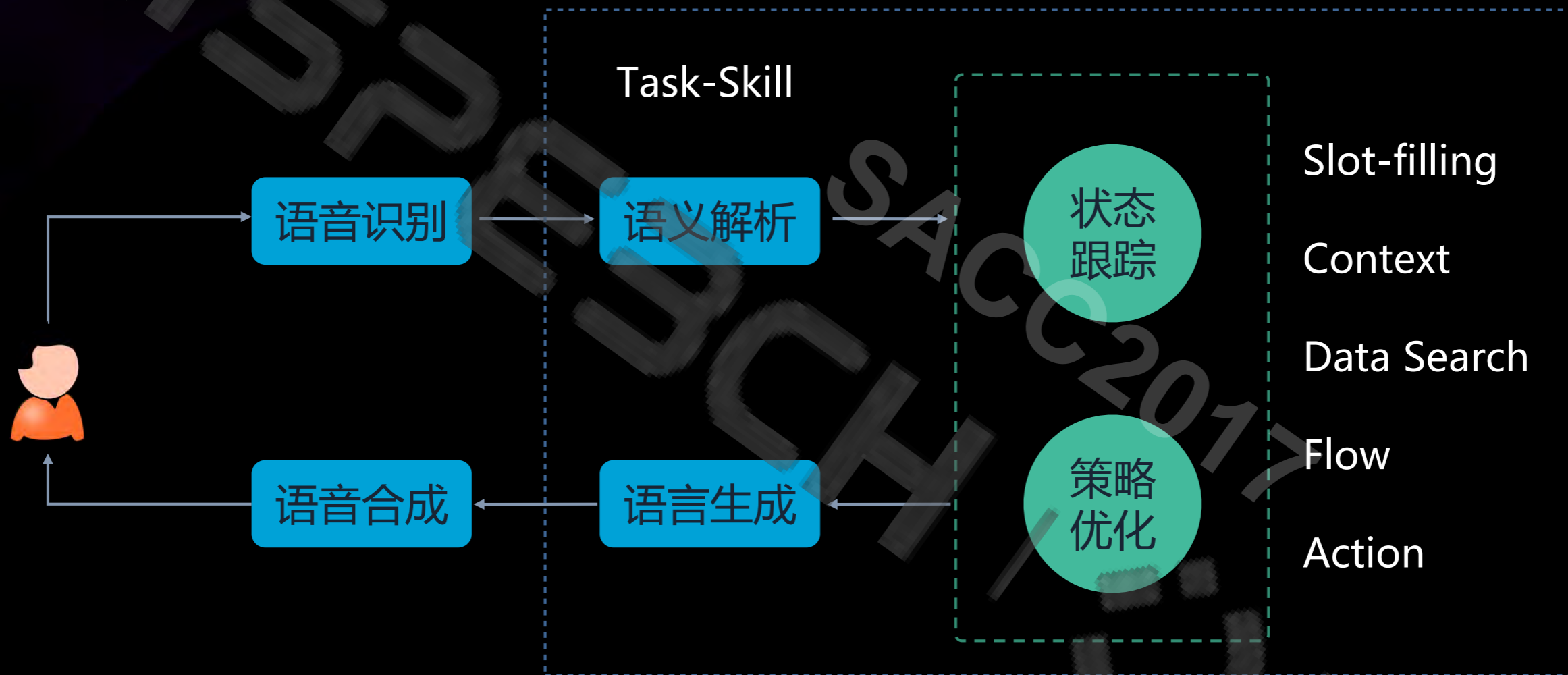
帮我导航到北京

把家里的空调打开

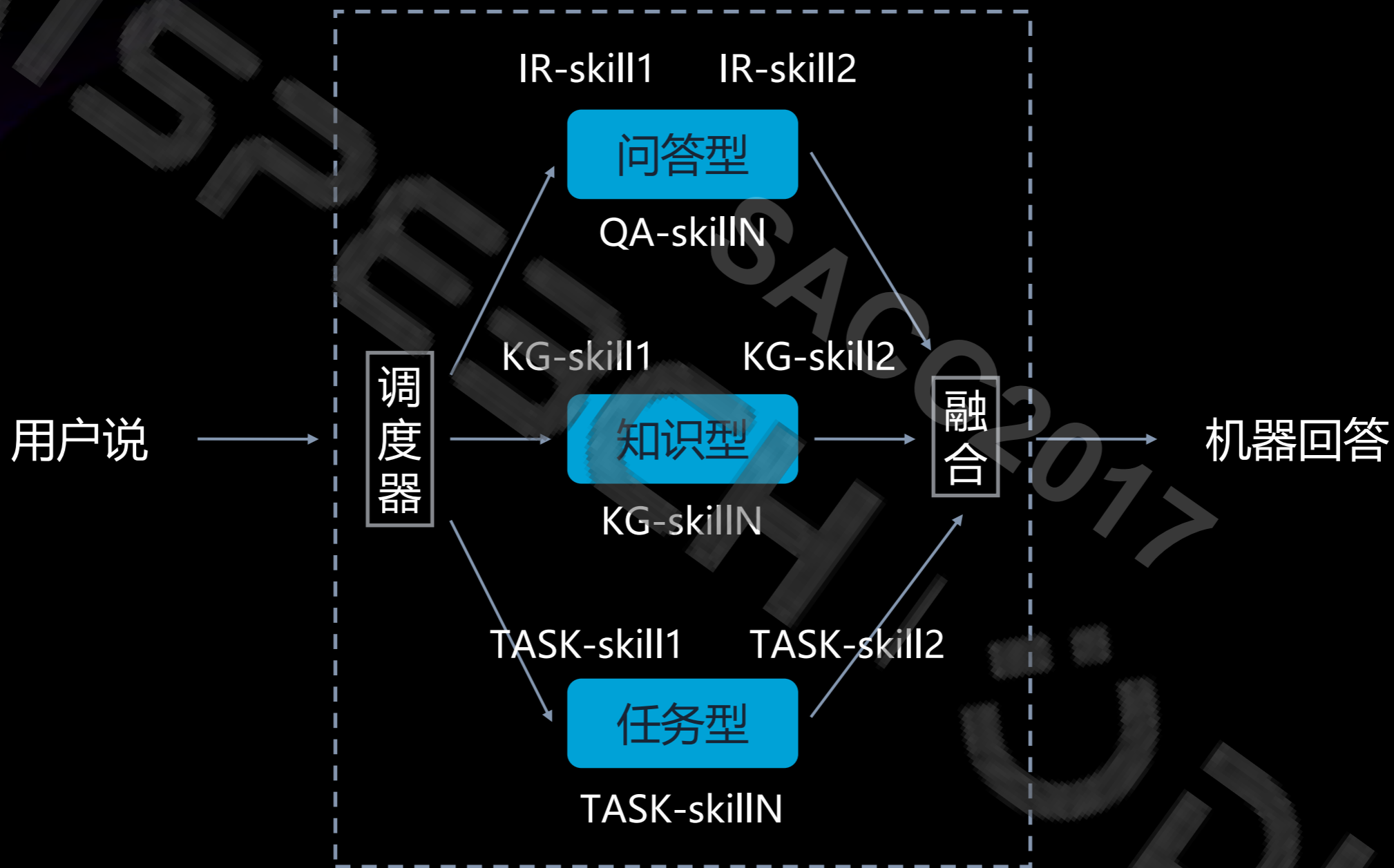
我要听岳云鹏的歌曲

北京今天天气怎么样?





面向任务的多轮对话

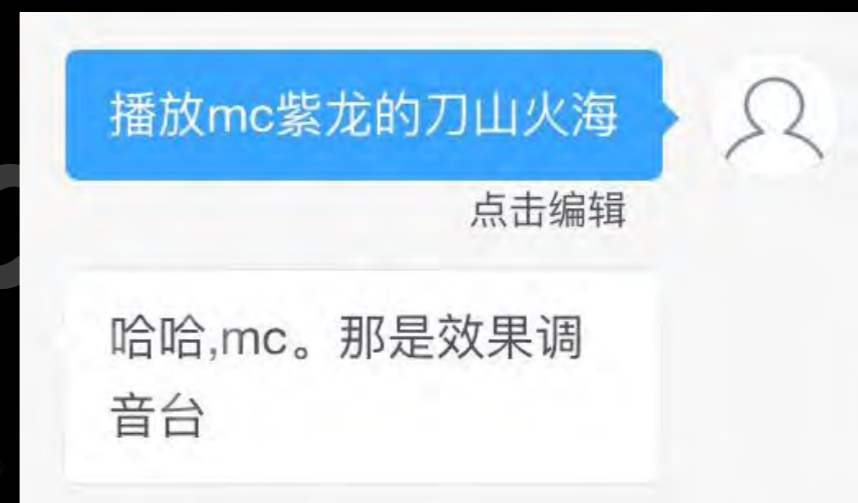


多引擎Skill架构

# PART 3

## 一些问题与思考

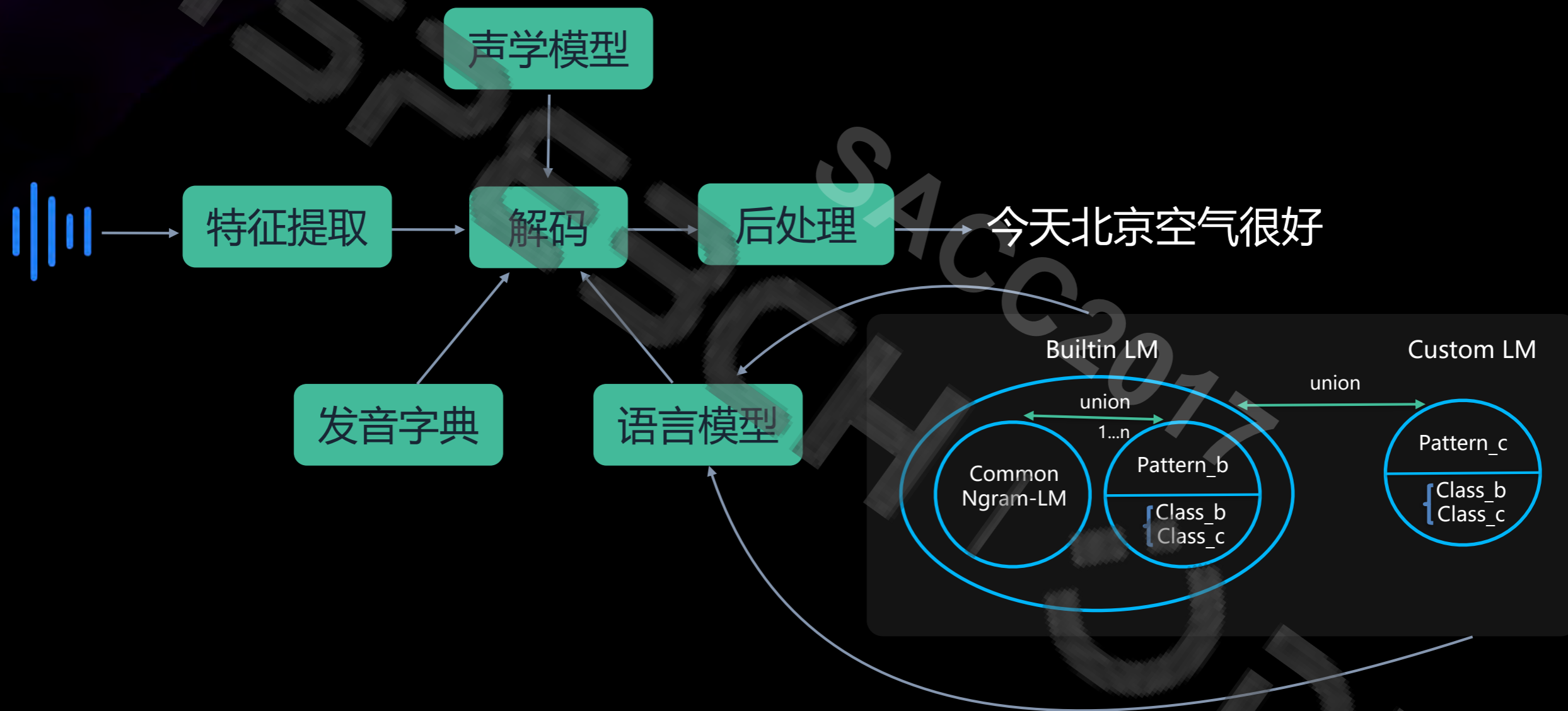
- 内容资源、语言模型与语义需要三者配合好
- 仅有识别或语义是不够的
- 识别与语义不宜分开



播放 “MC子龙” 的<刀山火海>

歌手名

歌曲名



内置语言模型 + 定制的Skill的语言模型



用Costa订杯咖啡

帮我订杯咖啡

喝咖啡要不要放糖

???

Costa Skill

Starbucks Skill

3W Coffee Skill

Coffee IR Skill

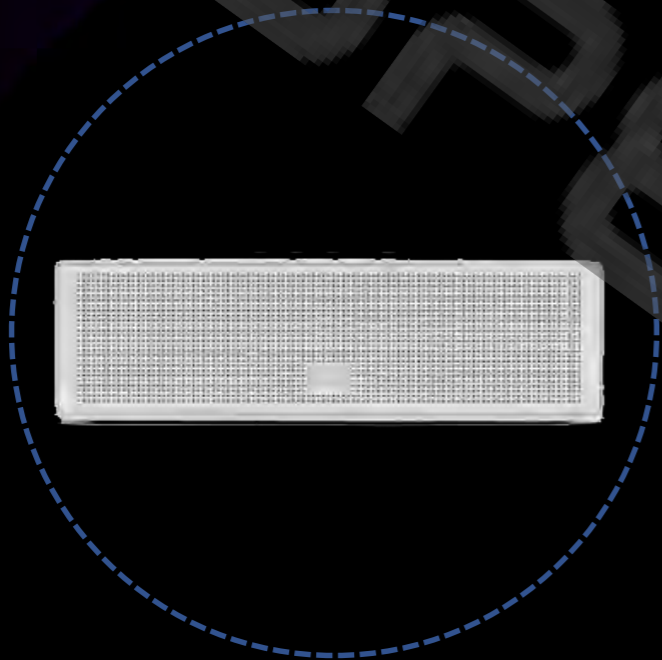
强说法

指名道姓

技能优先级

对话消歧

用户画像



GUI与VUI状态同步  
或无状态

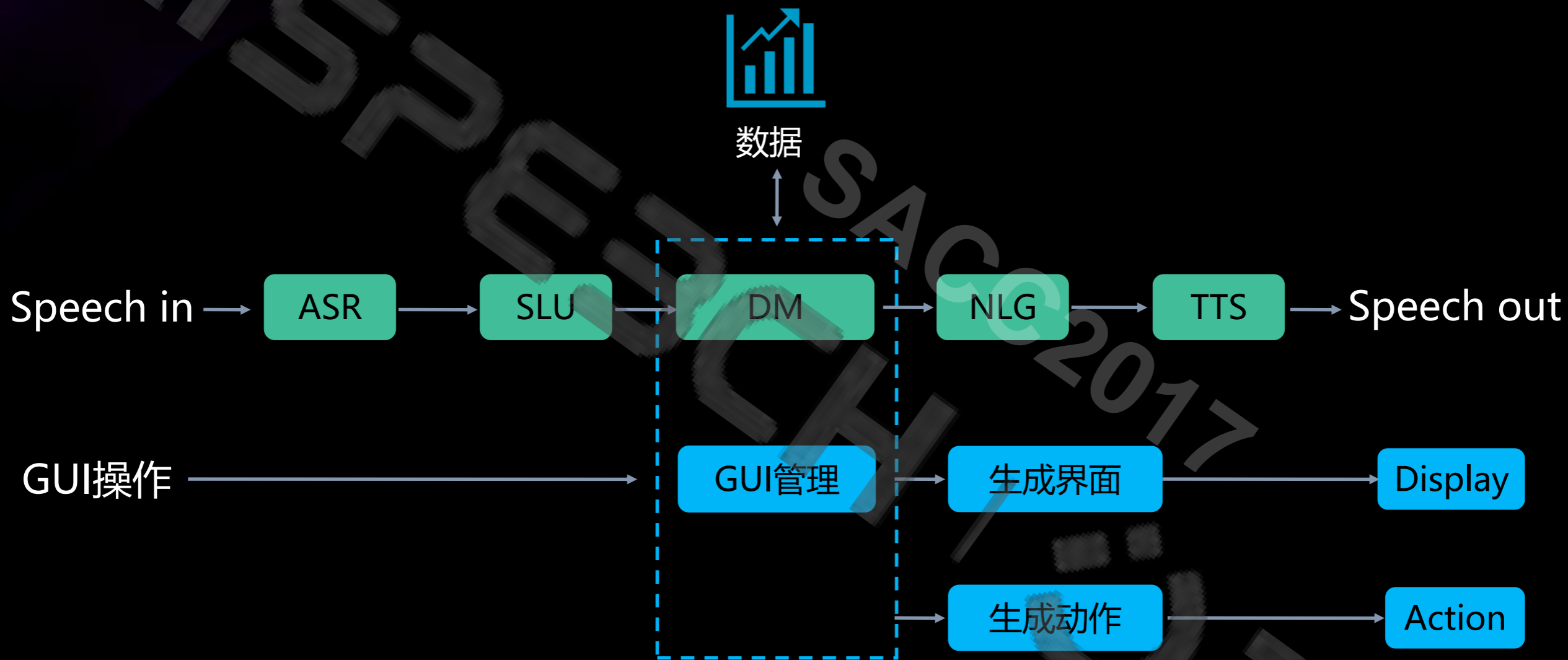
多模态交互  
VUI + GUI



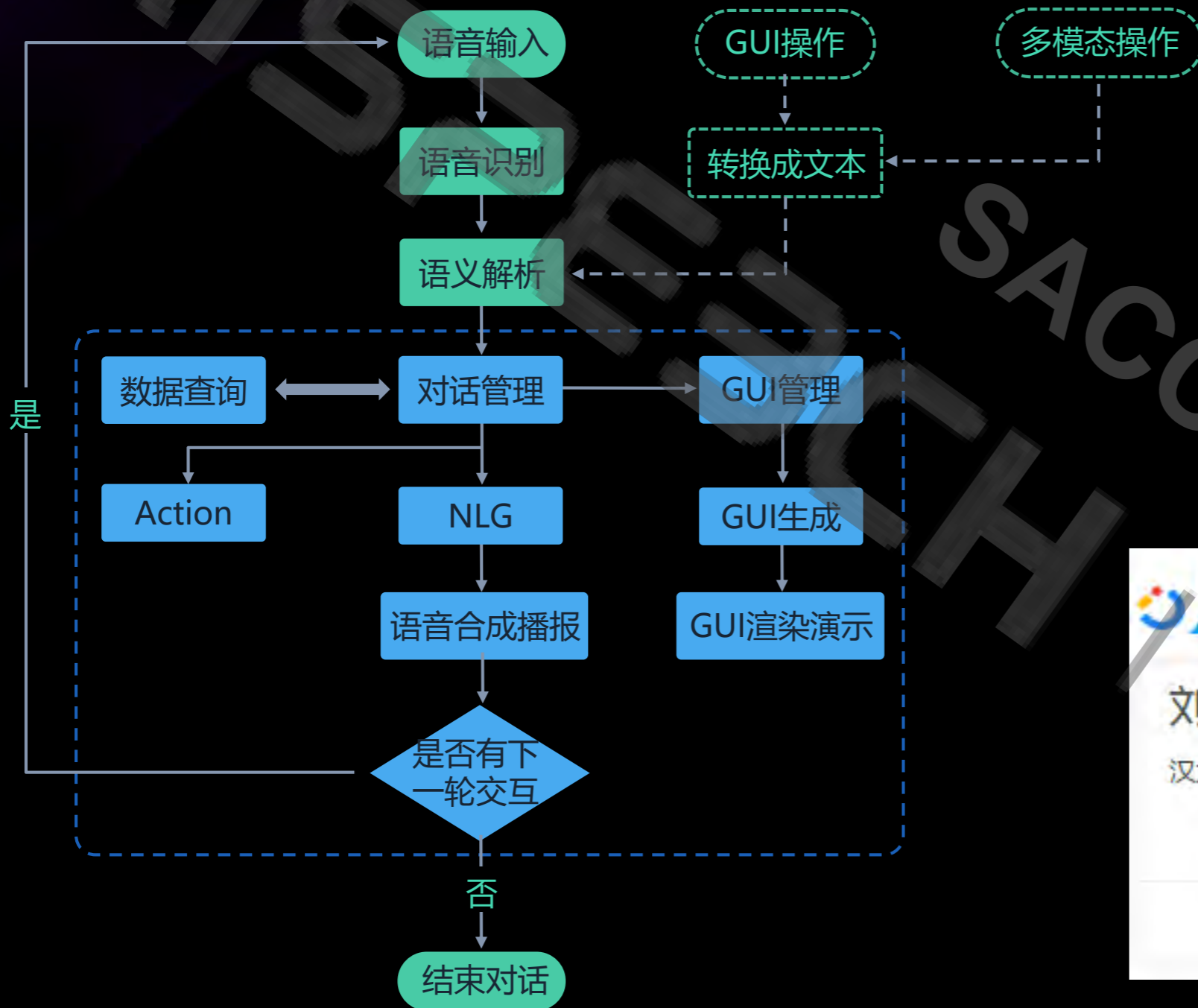
GUI辅助对话  
还是对话辅助GUI?



# Spoken Dialog System



Spoken Dialog System



### DUI控件

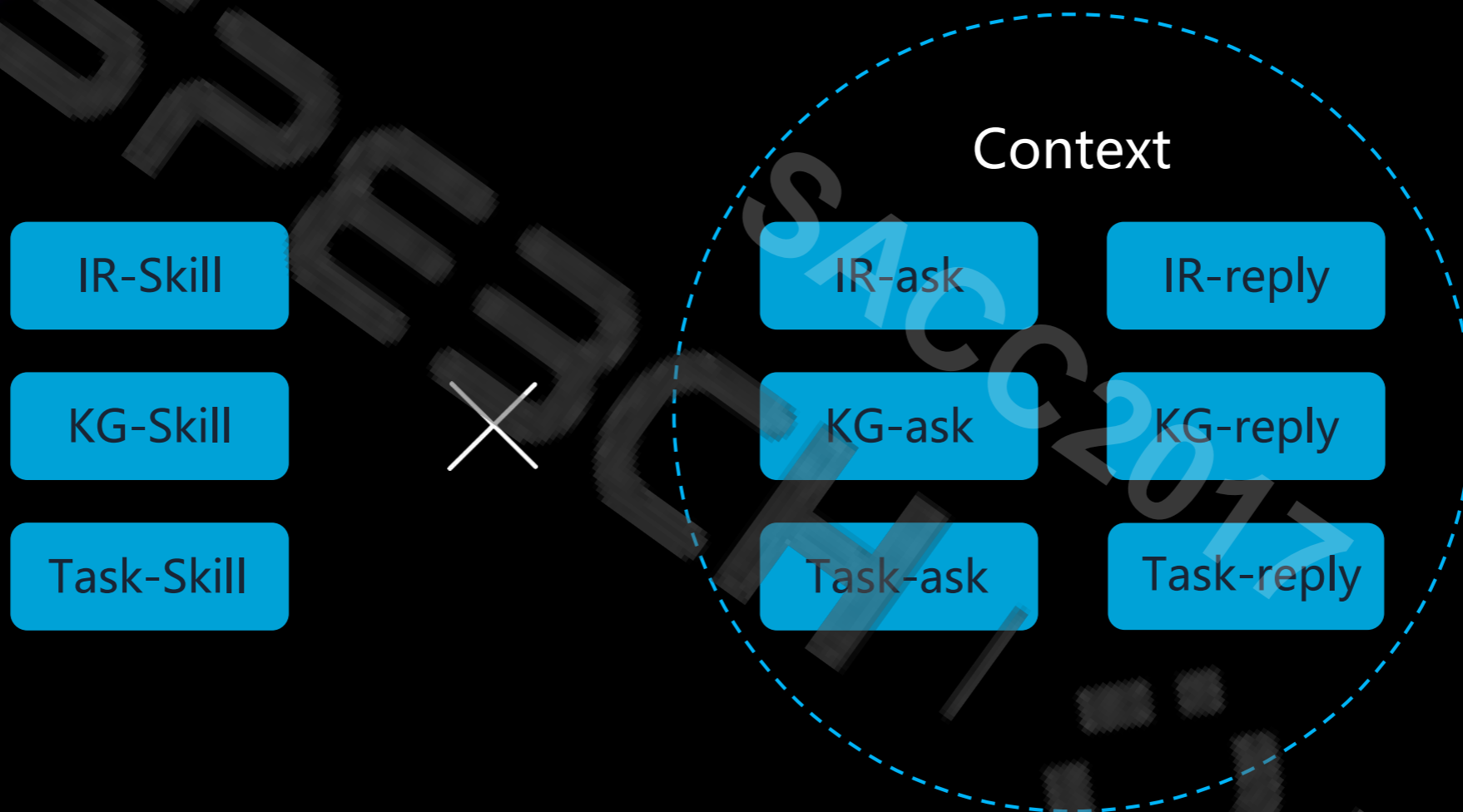
- 控件名称
- 控件类型
- 控件三要素
  1. Data
    - 数据格式
    - 数据源地址
  2. GUI
    - GUI布局



- 多引擎(IR,KG,Task)的融合
- 跨Skill的上下文
- 指代消解
- 知识推理???

给我来一首章子怡的老公的歌





指代消解的18种case, 引擎越多越复杂

## 无网情况下也可使用部分功能

- 并不是所有时候网络都是可靠的
- 需要云+端方案

打电话给张三



打电话

调大音量



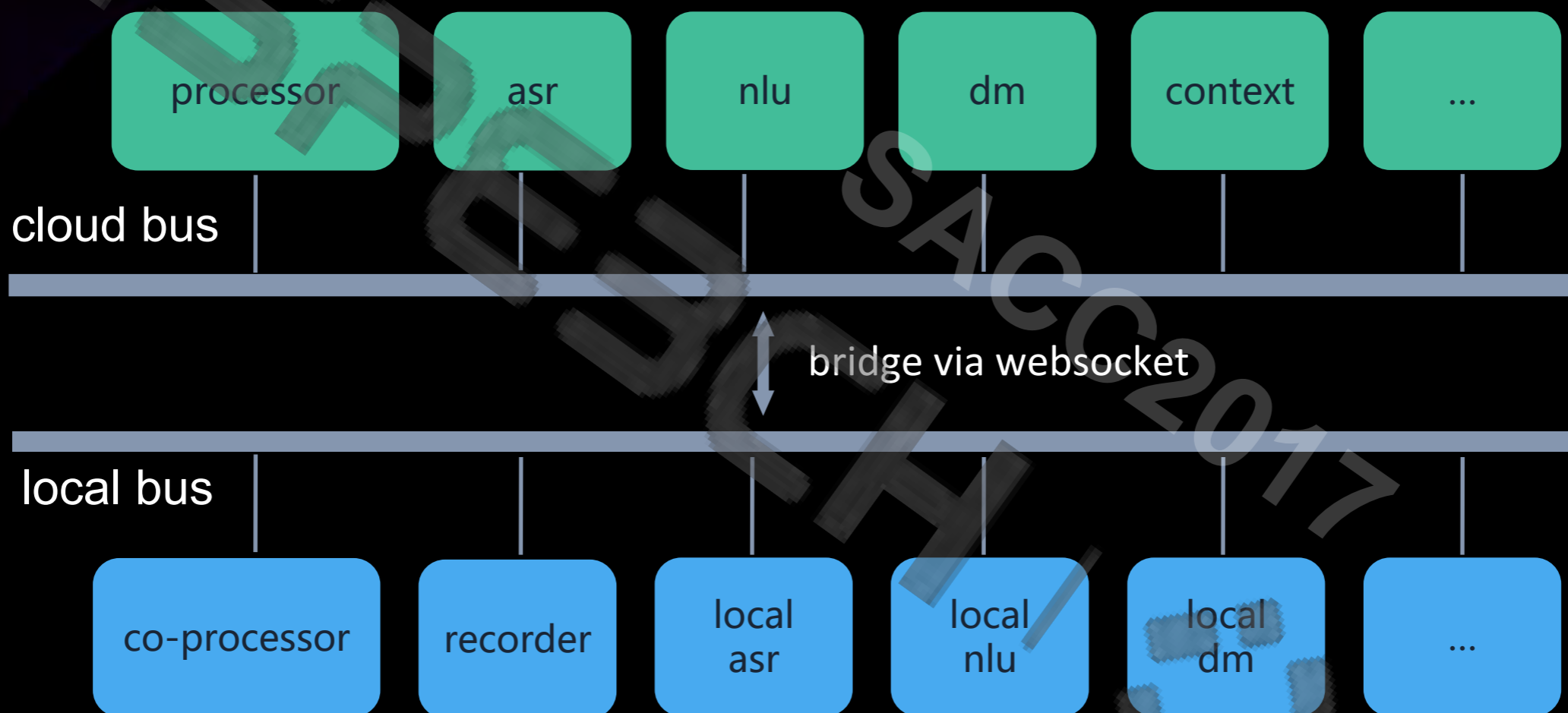
简单命令

导航到北京富力万丽酒店



导航





### 本地技能: 本地也可以运行ASR, SLU, DM

当网络弱或网络差时切换到本地技能, 只支持简单的不依赖服务端资源的技能

1. 为什么每次都要唤醒?
2. 为什么要等机器说完后人才可以说话?
3. 为什么要人主动和机器说话?
4. 功耗更低
5. 噪音环境下也可以交互
6. 机器实在搞不定怎么办?
7. 更智能的打断
8. 更智能的纠错
9. 更智能的知识推理
10. 更智能的指代消解
11. 更智能的技能调度
12. ...

# 思必驰核心技术介绍



# 从识别、理解到交互， 拥有全面的AI语音技术，全面拥抱对话智能



## 语音识别

- 实时云识别
- 大词汇识别
- 本地语音识别
- 抗噪及远场识别



## 语音合成

- 童音
- 名人合成音
- 标准男女声
- 个性化定制
- 高质量合成
- 超快速合成



## 语音识别++

- 语音唤醒
- 声纹识别
- 年龄识别
- 情绪识别



## 语义理解

- 电话短信
- 音乐电台
- 导航周边
- 天气日历
- 票务股票
- 设备控制等



## 智能对话

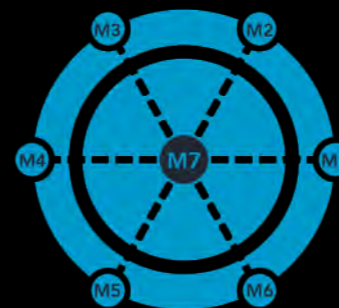
- 自由打断
- 智能纠错
- 渐进理解
- 任务对话
- 跟踪意图

# 全面的核心技术

## 识别

- 通用模型识别率在**97%**以上
- 提出了VDCNN模型, 在Aurora4噪音测试集上的识别性能达到了**7.09%**词错误率, 性能相对提高**10%**
- CTC模型, 降低**80%**搜索空间, 将识别速度提高**7倍以上**
- CPU+GPU混合架构
- 速度**100ms**以内

## 远场



- 6麦环形麦克风阵列, 3米识别率达**94%**

## 唤醒

### 定制唤醒

- 唤醒率 **95%**

### 通用唤醒

- 唤醒率 **92%**

## 语义

- 内置近百个领域, 覆盖车载, 电视, 故事机, 音箱, 机器人, APP等场景
- 支持亿级大词库, 解析速度提高20倍
- 支持定制说法自动调优
- 100ms以内

## 合成

- 标准男女声、可爱童声、孙悟空、周星星、志玲等十多种合成音色

## 问答

- 知识库 - 亿级百科知识

## 对话

- 多轮对话
- 跨领域上下文
- 纠正
- 指代消解
- 知识推理

# 和客户一起打磨产品



THANKS

MISSOURI STATE COLLEGE OF HUMANITIES