

# 360AI音箱云端测试之路

-基于case标注的测试系统

QA部.卫士驱动部 吴遐

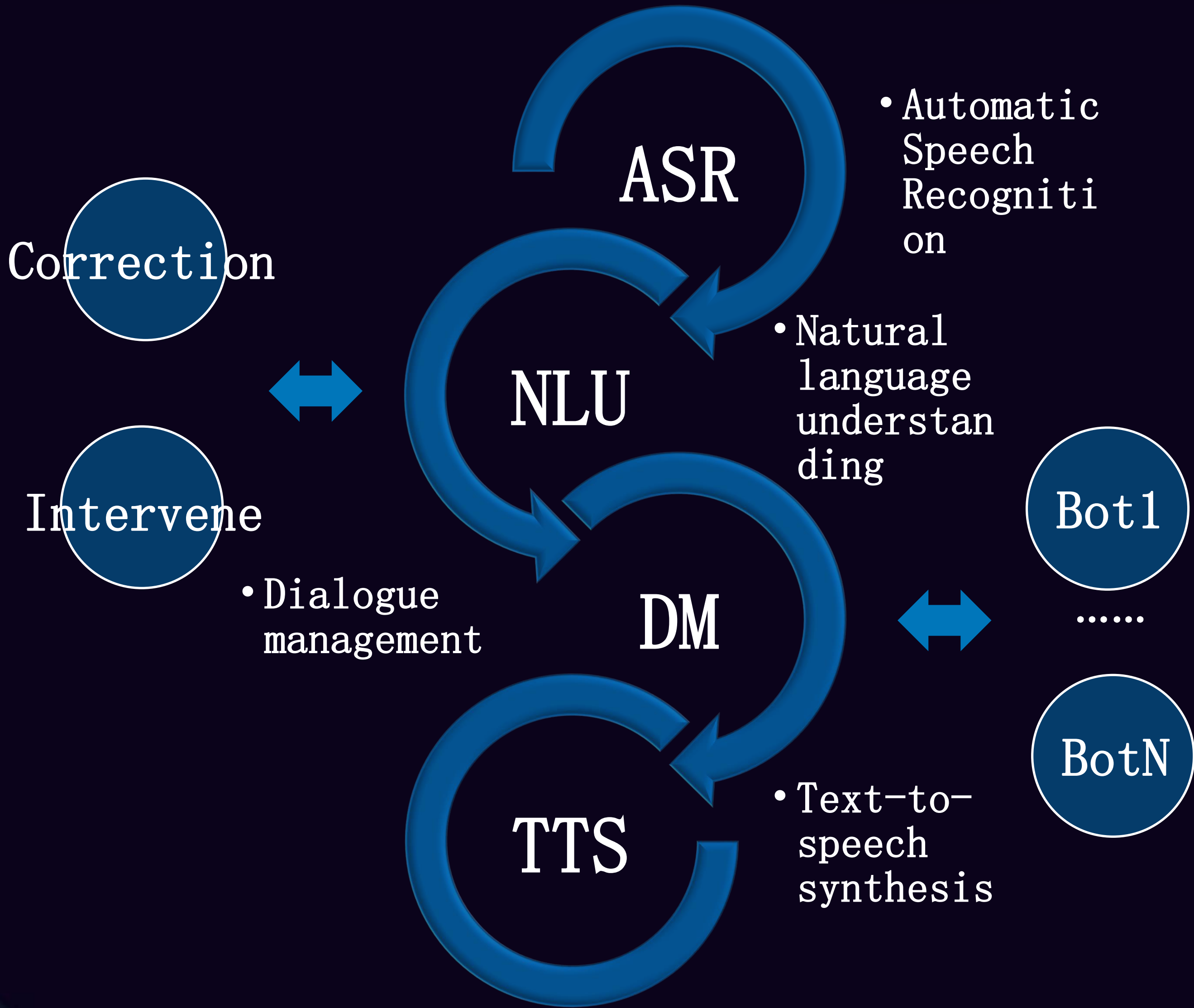
音箱云框架

case类型及扩充方案

case标注-提升自动化准确度

基于case的测试系统

展望







我要听刘德华的忘情水



agent: 音乐 intent: 播放 singer: 刘德华 song: 忘情水



找到资源: 刘德华的忘情水



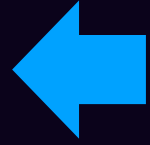
好的, 为您播放刘德华的忘情水



ASR: 音频转文本



NLU: 语义理解



BOT: 搜索对应资源



DM: 场景决策



TTS: 文本转音频

音箱云框架

case类型及扩充方案

case标注-提升自动化准确度

基于case的测试系统

展望

- 今天天气怎么样
- 为我播放一首好听的歌
- 我要听故事
- 打开扫地机器人
- 周杰伦
- .....



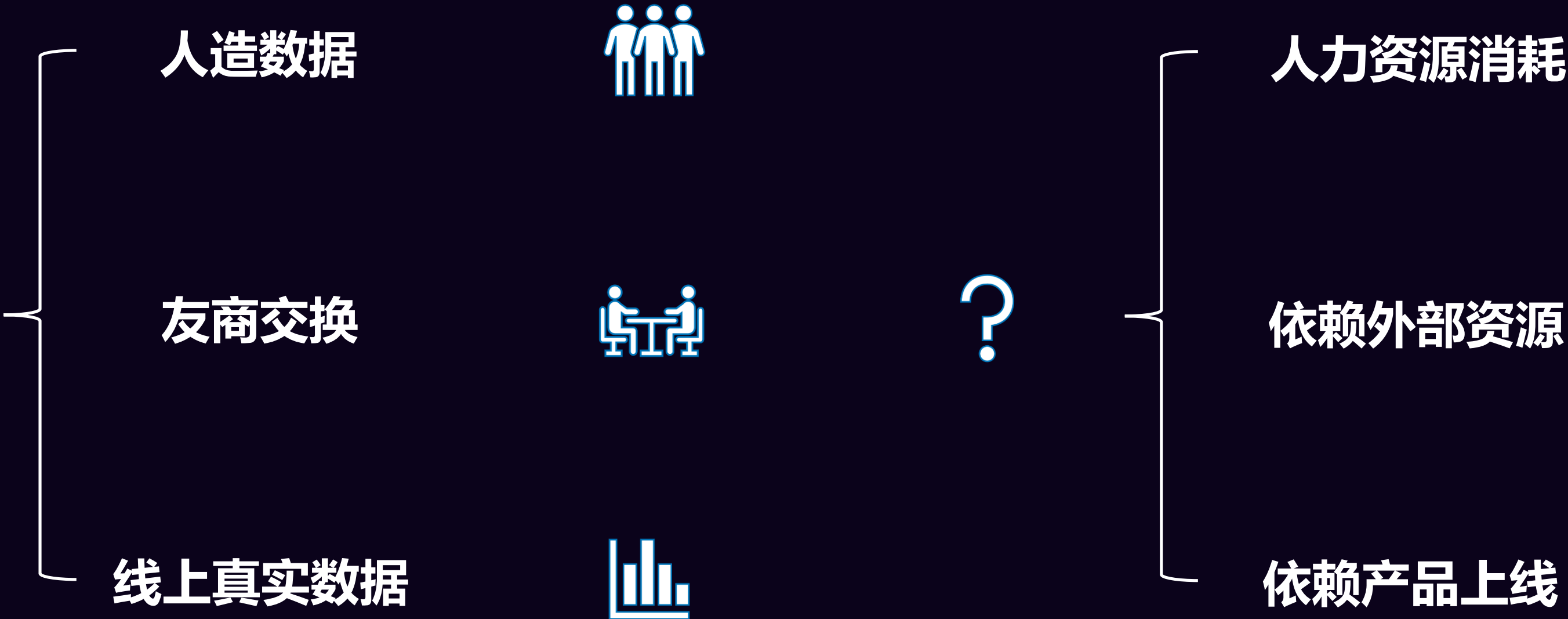
- 北京市今天，晴，.....
- 好的，为你播放稻香
- 为你推荐晚安妈妈
- 好的
- 周杰伦，1979年7月18日.....
- .....

query / case

answer

未标注case

case的来源





# 未标注case的优缺点

## 优点



- 来源多，数量大
- 无需人为处理case

## 缺点



- case结果无预期，需人工审核
- 结果单一，无法判断流程bug
- 各模块中间结果未知，排查问题难

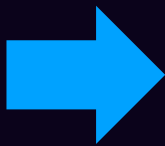


有标注case

input

output

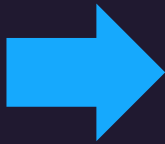
A. 播放刘德华的忘情水



A. 好的，为你播放刘德华的忘情水

需人工校验，效率低下

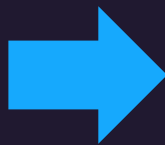
B. 播放刘德华的忘情水



B. agent: 音乐  
B. 好的，为你播放刘德华的忘情水  
B. 资源: 刘德华的忘情水

分类自动校验，效率有提升

C. 播放刘德华的忘情水



C. agent: 音乐  
C. intent: 播放  
C. singer: 刘德华  
C. song: 忘情水  
C. 资源: 刘德华的忘情水  
C. 好的，为你播放刘德华的忘情水

全自动校验

有标注case

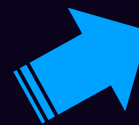
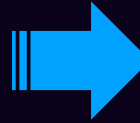
# 有标注case扩充

## case扩充方案一：人工标注

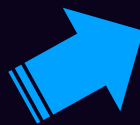
播放刘德华的忘情水



北京今天的天气



设置明天早上6点的闹钟



agent: 音乐  
intent: 播放  
singer: 刘德华, song: 忘情水  
资源: 刘德华的忘情水  
好的, 为你播放刘德华的忘情水

agent: 天气  
intent: 查天气  
location: 北京, date: 今天  
北京市今天, 晴, 5-12摄氏度, .....

agent: 闹钟  
intent: 设置闹钟  
time: 明天早上6点  
alarmid: 123456789  
已经为你设置了明天早上六点的闹钟

优点: 精准语义  
缺点: 人力资源, 慢

case扩充方案二：模板扩充

音乐资源：  
singer name: 刘德华  
song name: 忘情水

Template:  
播放<singer name>的<song name>  
我想听<singer name>的<song name>  
播放<singer name>的歌  
播放<song name>  
.....

<location>	<date>	
北京	今天	的天气
上海	明天	
广州	10月20日	

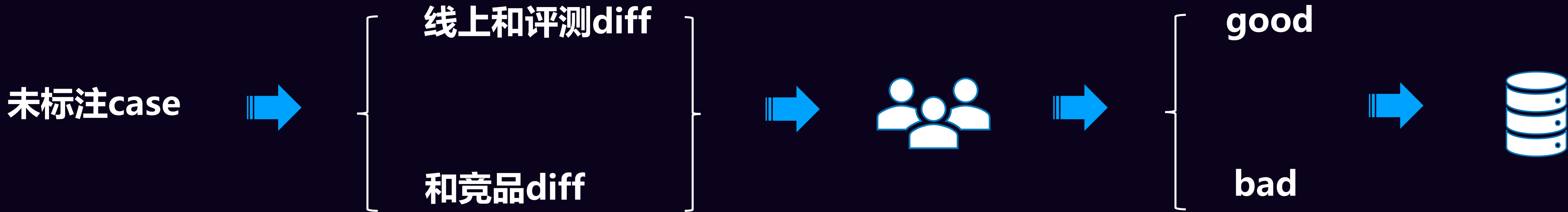
case:  
播放刘德华的忘情水  
我想听刘德华的忘情水  
播放刘德华的歌  
播放忘情水  
.....  
agent: 音乐  
Intent: 播放  
singer: <singer name>, song: <song name>  
好的, 为你播放<singer name>的<song name>

case:  
北京今天的天气  
北京明天的天  
北京10月20日天气  
.....  
agent: 天气  
Intent: 查天气预报  
location: <location>, date: <date>  
<location><date>, 晴, .....

优点: 自动标注, 量大  
缺点: 数据不规范, 歧义



case扩充方案三：评测case自动回收



线上结果

agent: 有声书  
intent: 播放  
video: 刘德华的忘情水  
资源: 刘德华的忘情水  
好的, 为你播放刘德华的忘情水

测试结果

agent: 音乐  
intent: 播放  
singer: 刘德华, song: 忘情水  
资源: 刘德华的忘情水  
好的, 为你播放刘德华的忘情水

good



优点: 自动标注+人工检查  
持续扩充  
缺点: diff量有限, 持续积累



音箱云框架

case类型及扩充方案

**case标注-提升自动化准确度**

基于case的测试系统

展望

## case标注-目标：全链路检查

case举例：播放牛德华的忘情水

### CORRECTION

- 播放刘德华的忘情水

### NLU

- agent: 音乐
- intent: 播放
- singer: 刘德华,  
song: 忘情水

### DM/BOT

- 资源: 刘德华唱的忘情水
- 来源: kugou

### 应答

- 好的，为你播放刘德华的忘情水

# 问题：自动化检测失败率高

## 场景

- 同一条case可能被多场景召回，但结果均为正确
- 举例：播放明月几时有。音乐？诗词？

## 槽位

- 多种槽位命中组合可能
- 举例：播放宝贝宝贝。儿歌名称？专辑名称？

## 资源

- 来源：多种来源可能，kugou, baidu
- 属性多样：资源名称、资源类别、资源id.....

## 应答

- 随机回复：好的/好呀
- 模板多样：为你播放刘德华的忘情水；为你播放忘情水；为你播放刘德华的歌； .....
- 服务异常检测不到：服务不稳定

## 多轮

- 单轮标注，需要考虑多轮场景

- 使用mongo做为底层存储：
  1. 数据结构清晰，多轮case下分多step;
  2. case属性增减灵活，结构任意修改，适合槽位变化;
- 多场景标注，综合评价：
- 槽位逻辑关系标注，匹配多种正确结果：

```
1 {
2     .....
3     "steps" : [
4         {
5             "test_query" : "播放明月几时有",
6             "or" : [
7                 {
8                     "AgentName" : "音乐",
9                     "IntentName" : "play",
10                    "songname" : "明月几时有",
11                    .....
12                },
13                {
14                    "AgentName" : "诗词",
15                    "IntentName" : "play",
16                    "name" : "明月几时有",
17                    .....
18                }
19            ],
20            .....
21        }
22    ],
23    .....
24 }
```



- 使用mongo做为底层存储:

1. 数据结构清晰, 多轮case下分多step;

2. case属性增减灵活, 结构任意修改, 适合槽位变化;

- 多场景标注, 综合评价:

- 槽位逻辑关系标注, 匹配多种正确结果:

```
1 {
2     .....
3     "steps" : [
4         {
5             "test_query" : "播放宝宝巴士主题曲",
6             "or" : [
7                 {
8                     "AgentName" : "儿童",
9                     "IntentName" : "play",
10                    "multianswer" : "album|video_song/video|video_song",
11                    "album" : "宝宝巴士",
12                    "video" : "宝宝巴士",
13                    "video_song" : "主题曲",
14                    .....
15                }
16            ]
17        }
18    ],
19    .....
20 }
```

- 资源属性相关标注，检查资源正确性；
- 多种文本检查逻辑：
  - 1.包含、不包含、空、非空等；
  - 2.支持槽位填充：为你播放{歌曲名}；
  - 3.异常回复检测：“对不起，服务异常……”；
- 多轮case：

```
1 {
2     .....
3     "steps" : [
4         {
5             "test_query" : "播放王菲和那英的相约1998",
6             "or" : [
7                 {
8                     "AgentName" : "音乐",
9                     "Intent" : "play",
10                    .....
11                    "bot_singername" : "王菲|那英",
12                    "bot_song_name" : "相约1998",
13                    "bot_source" : "kugou",
14                    "bot_resourceid" : "abcdefg",
15                    .....
16                }
17            ]
18        }
19    ],
20    .....
21 }
```

```
1 包含:
2     "text" : "王菲|那英|相约1998"
3     "text" : "播放|{singername}|{songname}"
4     .....
5 不包含:
6     "text" : "没有找到/还没有"
7     "text" : "服务不稳定"
8     .....
```

- 资源属性相关标注，检查资源正确性；
- 多种文本检查逻辑：
  - 1.包含、不包含、空、非空等；
  - 2.支持槽位填充：为你播放{歌曲名}；
  - 3.异常回复检测：“对不起，服务异常……”；
- 多轮case：

```
1 {
2     .....
3     "steps" : [
4         {
5             "test_query" : "播放音乐",
6             "or" : [
7                 {
8                     .....
9                     "AgentName" : "音乐",
10                    "IntentName" : "play",
11                    .....
12                }
13            ]
14        },
15        {
16            "test_query" : "换一个",
17            "or" : [
18                {
19                    .....
20                    "AgentName" : "音乐",
21                    "IntentName" : "sec",
22                    .....
23                }
24            ]
25        }
26    ],
27    .....
28 }
```

## case标注-目标：全链路检查

### 多轮case举例：1. 播放童话；2. 换一个

```
1 {
2     .....
3     "steps" : [
4         {
5             "test_query" : "播放童话",
6             "or" : [
7                 {
8                     "AgentName" : "音乐",
9                     "IntentName" : "play",
10                    "multianswer" : "songname/album"
11                    "songname" : "童话",
12                    "album" : "童话",
13                    "bot_singer" : "光良",
14                    "bot_song" : "童话",
15                    "bot_album" : "童话",
16                    "bot_source" : "kugou",
17                    "text" : "2好的|播放|{songname},1服务不稳定"
18                },
19                {
20                    "AgentName" : "故事",
21                    "IntentName" : "play",
22                    "multianswer" : "",
23                    "type" : "童话",
24                    "bot_tags" : "童话故事",
25                    "bot_source" : "ximalaya",
26                    "text" : "2好呀|即将播放|童话,1服务不稳定"
27                }
28            ]
29        },
```

```
30     {
31         "test_query" : "换一个",
32         "or" : [
33             {
34                 "AgentName" : "音乐",
35                 "IntentName" : "sec",
36                 "multianswer" : "",
37                 "songname" : "",
38                 "album" : "",
39                 "bot_singer" : "光良",
40                 "bot_song" : "",
41                 "bot_album" : "",
42                 "bot_source" : "kugou",
43                 "text" : "2好的,1服务不稳定"
44             },
45             {
46                 "AgentName" : "故事",
47                 "IntentName" : "sec",
48                 "multianswer" : "",
49                 "type" : "童话",
50                 "bot_tags" : "童话故事",
51                 "bot_source" : "ximalaya",
52                 "text" : "2好呀|为你播放,1服务不稳定"
53             }
54         ]
55     },
56     .....
57 }
58 }
```



音箱云框架

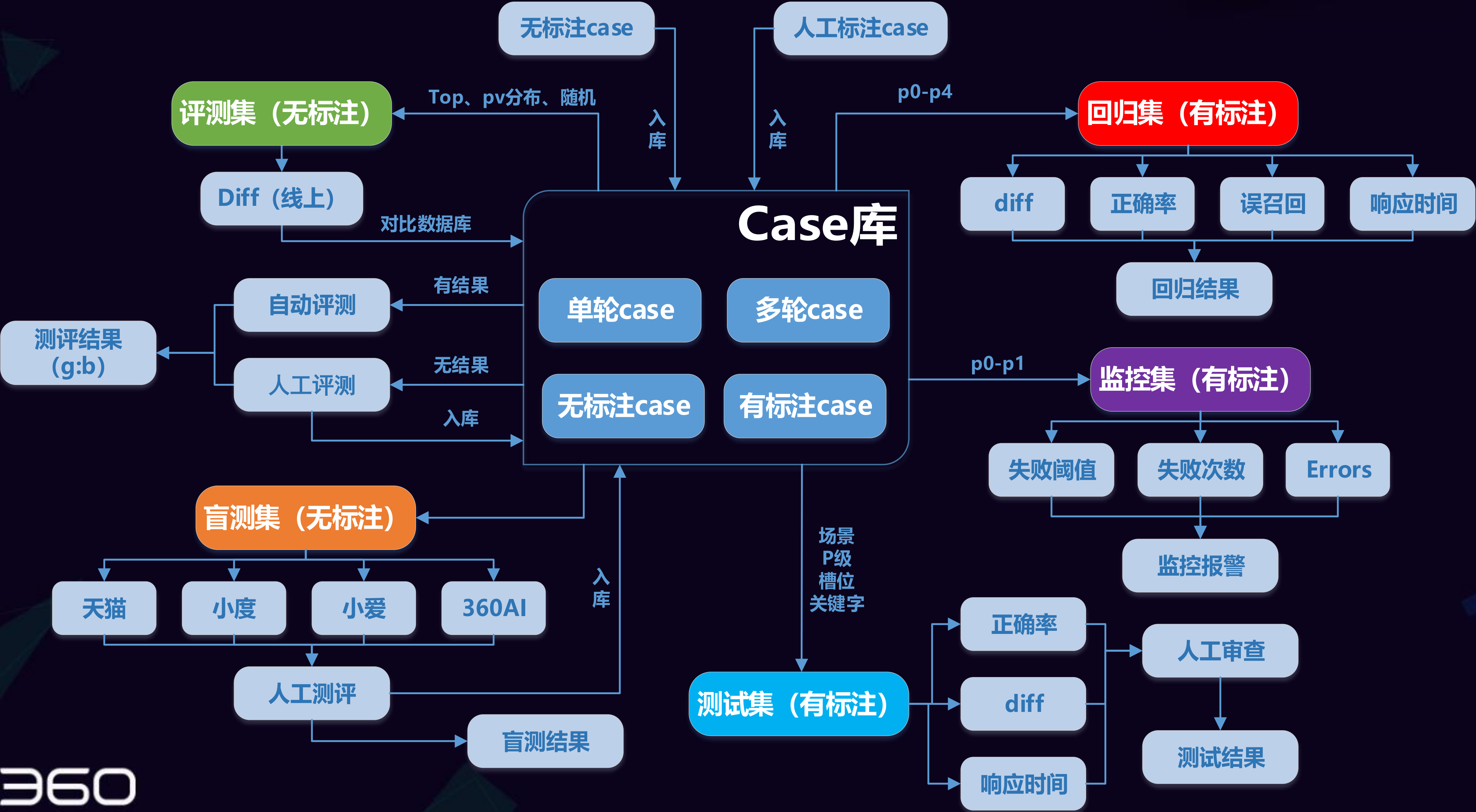
case类型及扩充方案

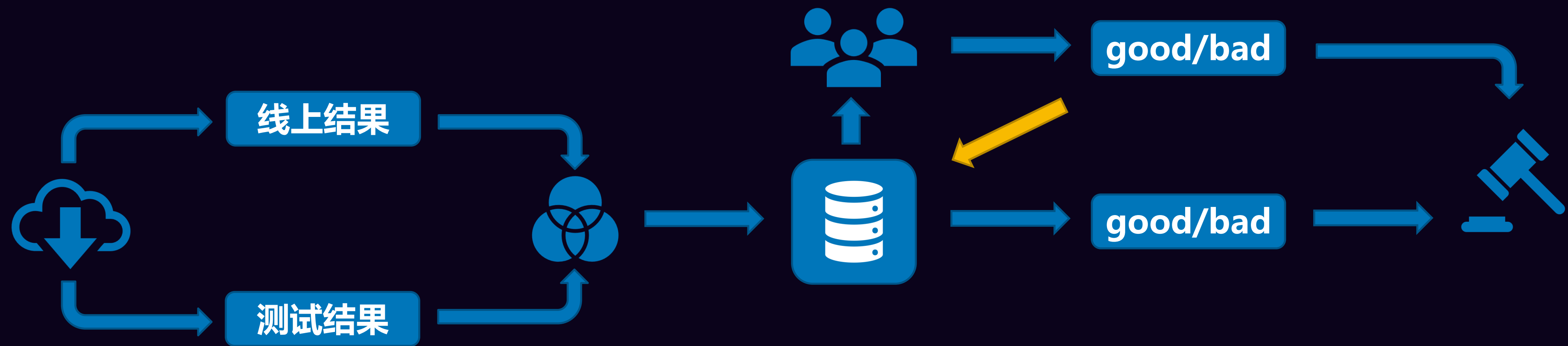
case标注-提升自动化准确度

**基于case的测试系统**

展望

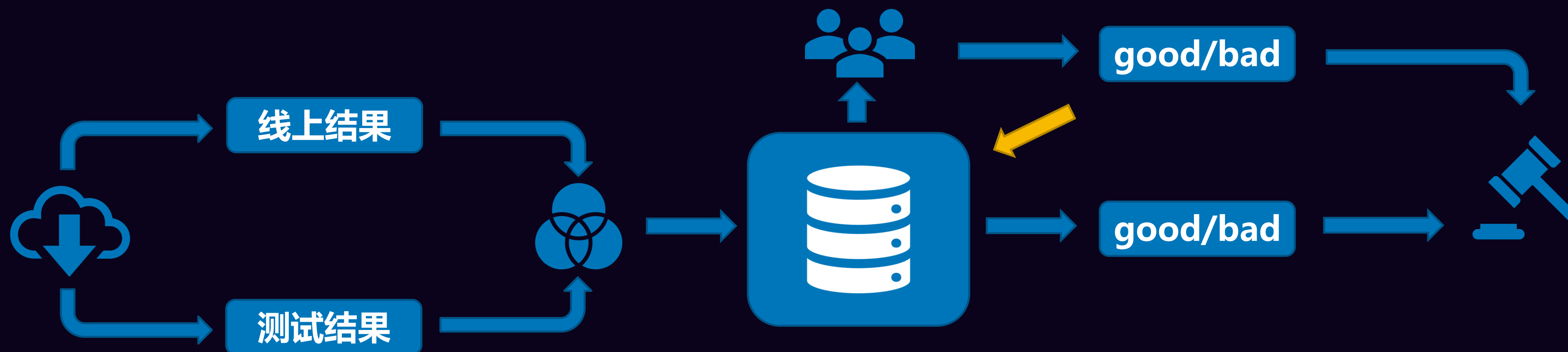




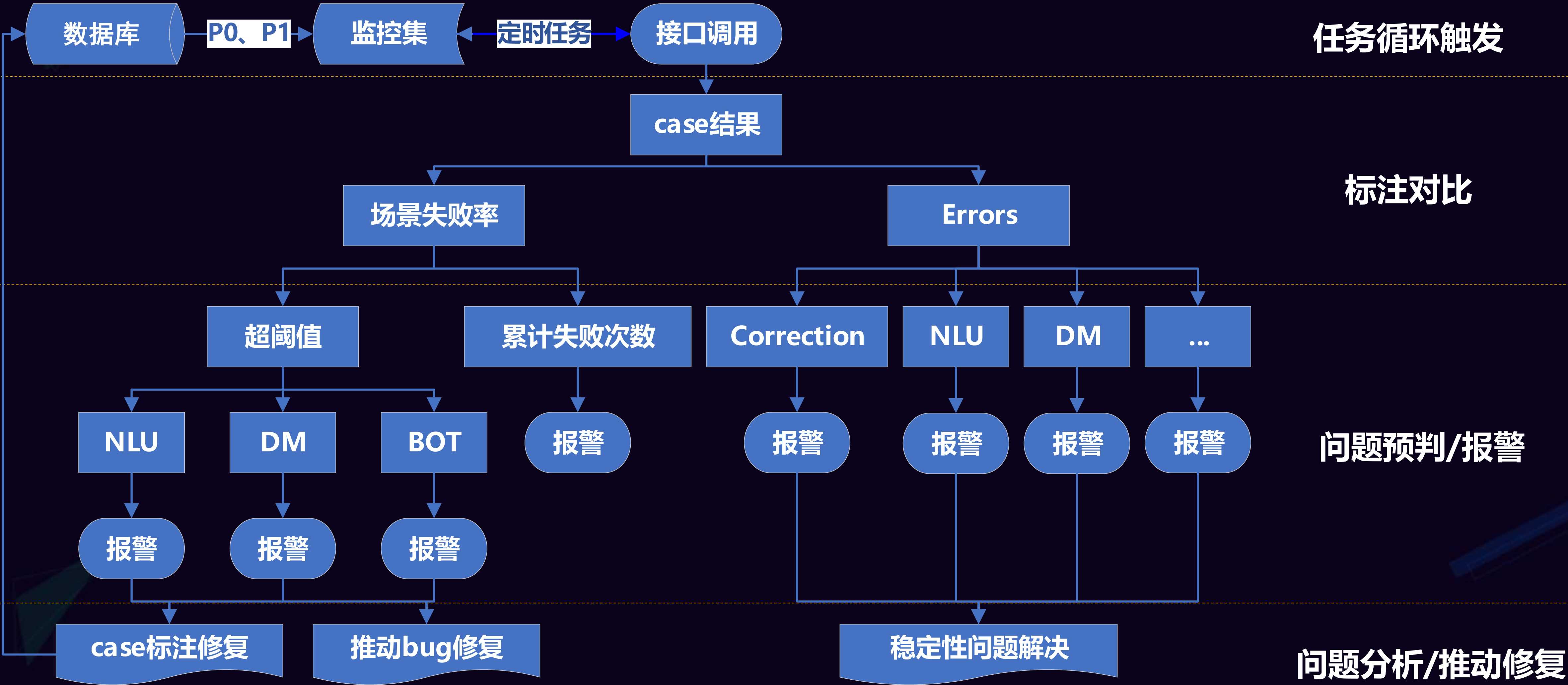


- 线上数据筛选：top、按pv分布、随机.....
- 自动评测+人工评测
- 人工评测结果入库，标注库累积增大
- 评测决策

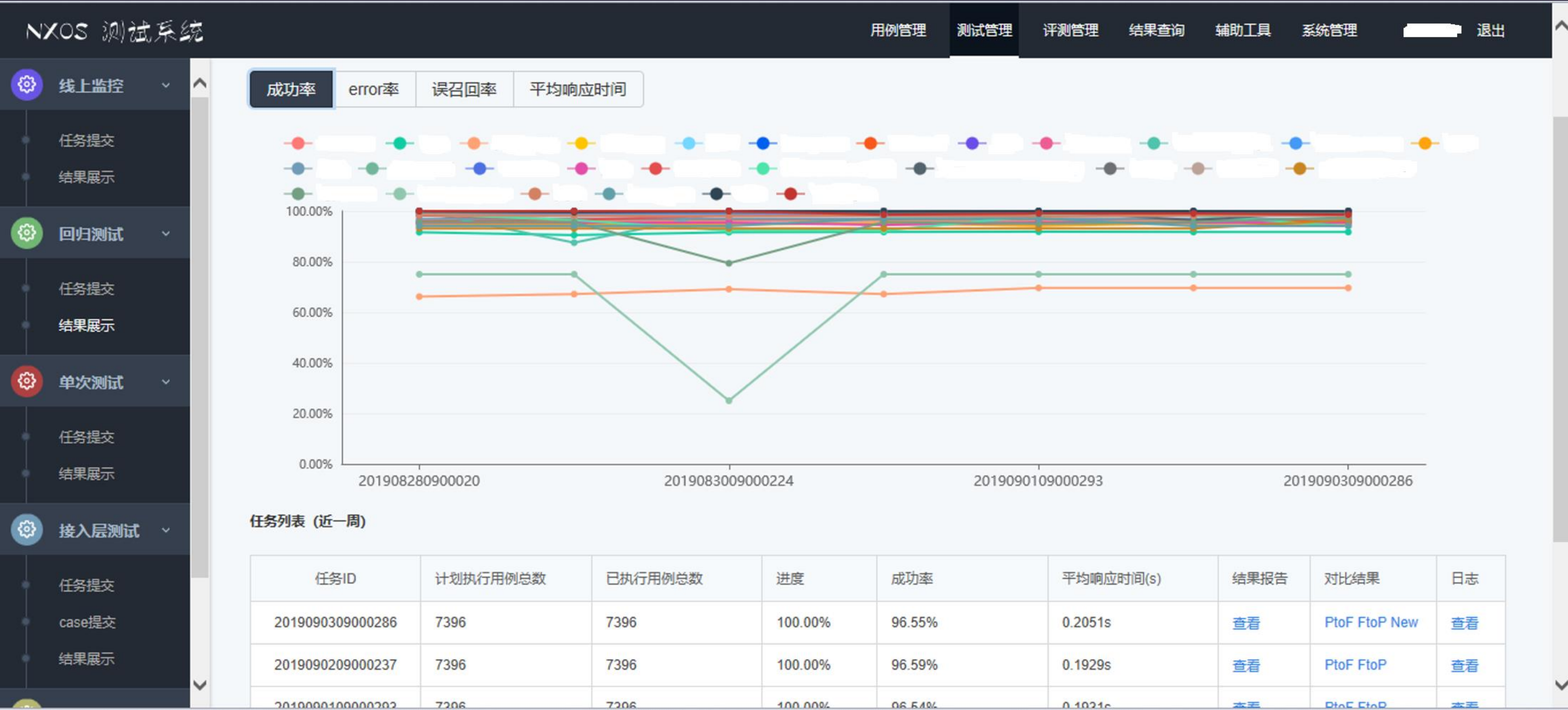




- 线上数据筛选：top、按pv分布、随机.....
- 自动评测+人工评测
- 人工评测结果入库，标注库累积增大
- 评测决策







音箱云框架

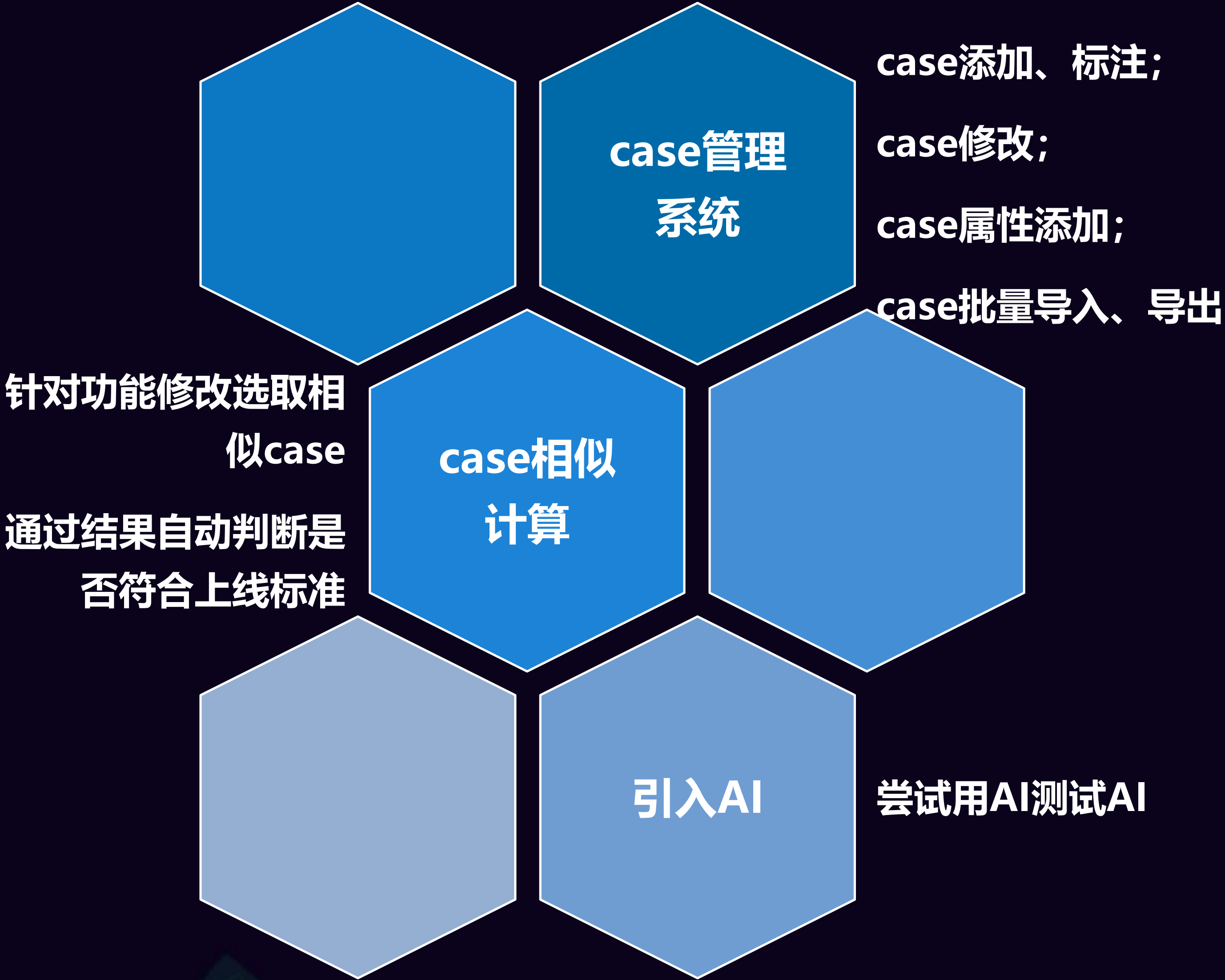
case类型及扩充方案

case标注-提升自动化准确度

基于case的测试系统

展望





Thank you!