



循环智能  
RECURRENT AI



# 智能质检

## 白皮书 (2022)

> 未来趋势：基于上下文语义的“非正则”质检



# 智能质检

## 白皮书 (2022)

> 未来趋势：基于上下文语义的“非正则”质检

# 目录（上）

## 一、未来趋势：“非正则” 5

- 为什么说“非正则”模式是未来趋势？
- 基于深度学习的“非正则”模式的优势
- “非正则”模式的技术原理
- “非正则”模式的底层逻辑
- 独有的“违规值排序”

## 二、新实践：双模质检 11

- “关键词+正则”模式的适配场景
- “非正则”模式的适配场景
- “双模”质检，各司其职

## 三、新实践：SOP流程质检 15

- 什么是“流程质检”？
- 实例：“单条件”流程质检
- 实例：“多条件”流程质检
- 监督不是目的

# 目录（下）

## 四、质检项优化 20

- 召回率和准确率的关系
- 负向质检项：通常“召回率”优先
- 正向质检项：通常“准确率”优先
- 实际场景更复杂
- 数字背后是“人心”

## 五、微信质检 26

- 企业微信需开通“会话存档”
- 微信质检的常见问题FAQ

## 六、保险行业的实践 29

- 传统机器质检效果不佳
- 实践一：从“关键词”升级到“一段话的语义”
- 实践二：引入自定义字段和复杂流程&逻辑配置
- 实践三：引入秘密武器“违规值”评分
- 保险质检的“三级火箭”



## 一、未来趋势：“非正则”

市面上有很多面向销售和客服人员的语音、文本质检系统，其他绝大部分产品使用的是基于“关键词+正则表达式”实现方式。由于漏检和错检问题严重，这类质检方式越来越难以满足企业的实际需求。

## 为什么说“非正则”模式是未来趋势？

很多企业拥有中大型电销或客服中心，每天与客户产生成千上万次沟通。沟通的质量，决定了企业的收入或客户满意度。企业监督业务人员的沟通质量和执行力的过程，被称为质检。

市面上有很多面向销售和客服人员的语音质检系统、文本质检系统，绝大部分产品实际使用的是基于“关键词+正则表达式”的机器质检系统。

这种方法的主要优点是部署和上手使用都比较快，主要缺点是存

在非常严重的漏检和错检情况。就像一个漏孔很大的筛子一样，难以满足企业对质检的需求越来越精细、对质检效率要求越来越高的发展趋势。

因此，在“关键词+正则表达式”之外，我们开始越来越多地为客户提供基于深度学习模型的“非正则”方案，并且在实际使用中能够多发现 2~10 倍的问题。

对于企业而言，采用“非正则”方案意味着他们可以全面地提升服务质量或者实现合规升级。

(.\*)

“关键词和正则的方法，难以满足企业对质检的需求越来越精细、对质检效率要求越来越高的发展趋势。”

## 基于深度学习的“非正则”模式的优势

语音和文本质检的主要任务是找出业务人员“说了什么不该说的内容”或者“该说的什么内容没有说”。前者通常被称为负向质检；后者通常被称为正向质检，企业可以设置为“做得不好就减分”的惩罚制，也可以设置为“做得好就加分”的奖励制。

传统基于“关键词+正则表达式”的产品做质检，所遇到的最主要问题是“找不全”，通常会漏掉很多不合格、不合规之处，导致质检效率大打折扣。

看一个实际对比的例子。某互联网公司的基础质检项“服务态度问题”，在我们的实际应用中：使用传统“关键词+正则”方案，一天的数据中能找出 13 条，100% 是正确的；使用新的“非正则”方案，能找出 134 条，其中 72% 是对的。从最终结果看，新的“非正则”方案多找出

了8倍的问题。

再看一个实际的例子。“恐吓威胁”是贷后资产管理领域的基础质检项，即催收员不允许在电话里“恐吓威胁”债务人。在我们的实际应用中：使用传统“关键词+正则”方案，四天的数据中能找出 316 条，其中 55% 是正确的；使用“非正则”方案，能找出 2203 条，其中 72% 是对的。从最终正确的条数来看，174条对比 1596条，新的“非正则”方案能多找出 9 倍的风险。

原因其实很简单。如果使用基于“关键词+正则”的方案，方法是用关键词的组合来涵盖每个质检项的不同表达方式——但是你可以写10个关键词，100个关键词，却永远不可能穷尽，因为语言的表达方式是非常多样的、千变万化的，必须通过整个句子的上下文语义才能做出更准确的判断。

## “非正则”模式的技术原理

非正则方案，目标是训练一个深度学习算法模型，使之能够判断关键词未覆盖的句子是否命中了质检项。我们以另一个贷后资产管理领域常见的质检项“暴露客户隐私”为例。从标注到训练模型，再到最后上线使用，“非正则”方案大致分为三个步骤。

第一步，使用我们的“标注工厂”产品，通过人工的方式，将“暴露客户隐私”的句子标记为“正例”，将不是“暴露客户隐私”的句子标记为反例。

第二步，将一定规模的经过标注的正例和反例都“喂”给训练器，让训练器学习到一个算法模型，这

个算法就能用来判断新句子是不是涉嫌暴露客户隐私。

第三步，在质检产品中，系统就可以标记出所有命中“暴露客户隐私”语义点的句子，复检员可以快速定位到该质检项所处的位置，迅速进行核实。此外，复检员每一次复检的操作，都相当于对算法模型进行了一次反馈，会帮助算法模型变得更准。

最终，比“关键词+正则”方案多找出数倍的不合格、不合规之处，达到召回率（找的全）、准确率（找得准）均在 80% 以上的效果。



## “非正则”模式的底层逻辑

从底层逻辑上看，基于“关键词+正则”的方案是字符级别的，并不关心句子的语义，而基于深度学习模型的“非正则”方案是句子级别的，非常关心句子上下文的逻辑和语义。两者并不在同一个维度。未来“关键词+正则”方案越来越难当大任，而“非正则”的方案会逐步成为主流。

总结循环智能在金融、房产、汽车等不同行业标杆客户的服务经验，我们发现质检项与违规数的关系也存在“二八法则”——20%的质检项贡献了80%的违规数，所以将最常见质检项升级到“非正则”方案，即可为整个业务带来显著的效果提升。

同时，我们也应该了解到，随着自然语言处理领域新技术的突破，从字符级“关键词+正则”方案，向句子级“非正则”方案转换的速度正在加快。过去两年，自然语言处理领域迎来了繁荣时期。Google 发表于 2018 年的 BERT 模型，为行业带

来了全新的技术思路，具有里程碑意义。2019年6月，作为 BERT 模型的一种重要的改进方案，XLNet 模型在 20 个标准任务集上超过 BERT，并且在 18 个标准任务集上取得 state of the art 成果，包括机器问答、自然语言推断、情感分析和文档排序等。

XLNet 模型由循环智能联合创始人杨植麟博士（第一作者），与谷歌大脑、卡内基梅隆大学共同推出。该模型具备编码超长序列的能力——简单理解就是可以更好地理解长句子。

# XLNet

20 个人工智能标准任务集超过 BERT

18 个人工智能标准任务集取得 state of the art

## 独有的“违规值排序”

在负向质检——找出业务员“说了什么不该说的”——场景下，我们的贷后资产管理（催收）客户，希望尽可能找出所有违规对话。因为每个漏掉的对话，都可能给企业带来风险。

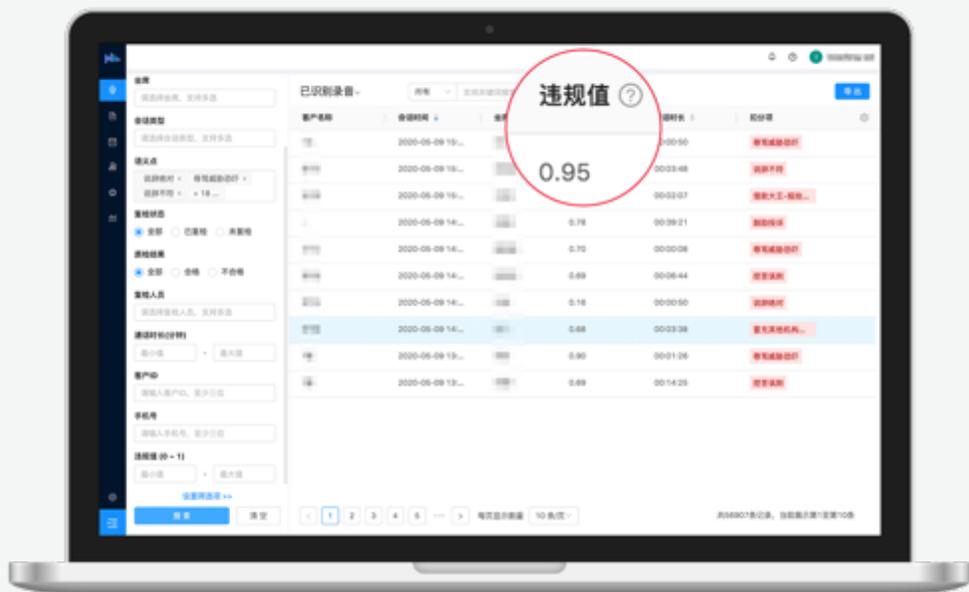
有些企业的人工复检员比较多，他们在复检过所有机器质检的结果之后，仍有剩余精力，这时就可以使用我们提供的一项独有功能：违规值排序。

简单来说，尽管大部分对话没有命中任何质检项，但我们也会通

过算法模型给所有对话打个分，看它违规的可能性是多少。算法是通过学习那些违规对话的特征做到的。

我们的一家客户，在“机器质检+人工复检”的工作流程之外，还会使用违规值排序这项功能进行抽检——从高违规值的对话中抽取一些进行检查。

结果发现，借助违规值排序的人工抽检，能比之前纯人工的完全随机的抽检，可以多发现2~3倍的违规对话。





## 二、新实践：双模质检

随着自然语言处理（NLP）领域技术的发展，以“非正则”质检模式为主、“关键词+正则”模式为辅的方案已成为未来发展趋势。两种模式将长期共存，因为它们各有其更擅长做的事情、更适配的场景。

## “关键词+正则”模式的适配场景

在实际应用中，“关键词+正则”作为一种基础质检模式，越来越难以满足企业在质检效果和效率上的精细化需求。但是，这种模式也有它的优势：上手快。

当企业提出一个新的、此前从未用过的质检项时，质检项的标准尚未完全确定，因此可以用“关键词+正则”模式先跑起来，快速进行探索和迭代。其后根据初步探索和迭代的结果，再判断是否可以升级到基于深度学习技术的“非正则”模式。这是适配“关键词+正则”模式的第一种场景。

第二种场景：当一个质检项命中的目标通话量比较少，只有几百甚至几十条，就无法产生足够的“正例”给机器学习模型进行训练，只能继续采用“关键词+正则”模式。这是一种被动场景，也是比较常见的场景。

还有一种场景比较特殊：有些质检项命中的目标通话比较多，原

则上可以用来训练机器学习模型，但是因为“关键词+正则”已经得出不错的结果，既找得全（术语叫召回率高），又找得准（术语叫准确率高），两个值都超过 90%，那么暂时就不迫切需要升级到“非正则”质检模式了。比如很多企业需要的正向质检项——“礼貌问候”，因为可以穷举出大部分“礼貌问候”的用词，用“关键词+正则”模式就能得到双 90%的结果。

不过，这种场景非常少见。大部分情况下，“关键词+正则”质检模式相比基于深度学习技术的“非正则”质检模式，在找全率和找准率上有很大差距。

### 关键词+正则

<b>优劣势</b>	<ul style="list-style-type: none"> <li><span style="color: green;">✔</span> 上手快</li> <li><span style="color: red;">✘</span> 规则代码的后期维护成本高</li> <li><span style="color: red;">✘</span> 找不全(大部分时候)</li> <li><span style="color: red;">✘</span> 找不准(大部分时候)</li> </ul>
<b>适配场景</b>	<ul style="list-style-type: none"> <li><span style="color: orange;">👍</span> 全新质检项，标准不完善，可快速上手</li> <li><span style="color: orange;">👍</span> 质检项产生的目标通话量较少，难训练模型</li> </ul>

## “非正则”模式的适配场景

在实际应用中，传统“关键词+正则”模式，很难找得全和找得准目标通话，大大影响工作效率——如果找不全，就意味着会遗漏很多目标通话；如果找不准，就意味着在人工复检时会浪费大量人力。“非正则”质检模式，可更好地解决上述痛点。

考虑到在实际使用中，质检项与命中的目标通话量之间的关系也存在“二八法则”——20%的质检项贡献了80%的质检量，所以将质检量大的少数质检项升级到“非正则”模式，往往可以大幅提升整个质检任务的找全率和找准率。这是“非正则”质检模式主要的适配场景。

第二种场景：某些质检项，虽然从某一家企业的角度看，所命中的目标通话量不算大，但是这个质检项是整个行业中较为成熟的、通用的质检项，其他企业也都在用，那么就可以采用基于深度学习技术的“非正则”模式进行模型训练。因为其他企业都可以比较快速的复用，所以“非正则”模式带来的收益就更高。

与“关键词+正则”质检模式相比，“非正则”模式，不再需要既懂业务又懂正则的稀缺人才编写规则和迭代规则，只需要普通人快速进行数据标注即可训练算法模型，而且通常能得到更好的效果。

### 非正则(深度学习)

#### 优劣势

- ✓ 找得更全(大部分时候)
- ✓ 找得更准(大部分时候)
- ✓ 质检项生产速度更快
- ✗ 若产生的目标通话量少,则无法训练模型

#### 适配场景

- 👍 产生的目标通话量够多的质检项(绝大部分)
- 👍 成熟的、行业通用的质检项

*“非正则模式，只需要普通人快速进行数据标注即可训练算法模型。”*

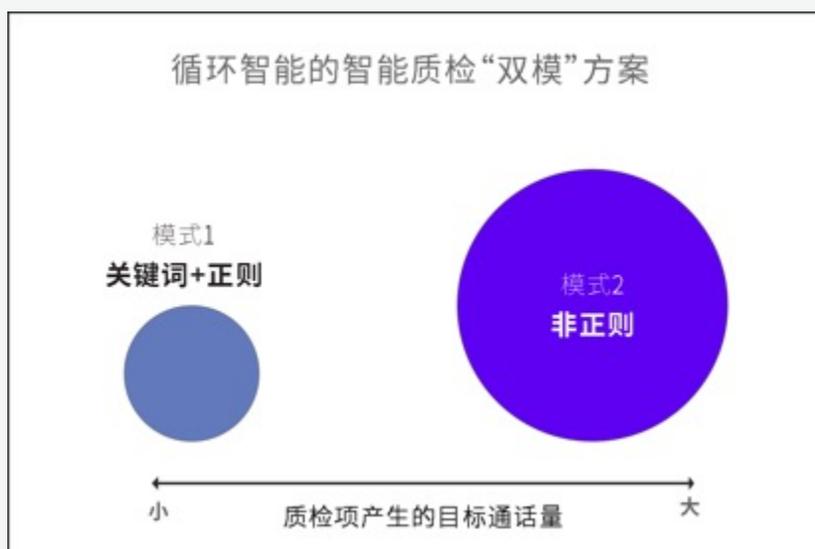
## “双模”质检，各司其职

当前阶段，质检项应该采取哪种模式，主要取决于该质检项产生的目标通话量大小——通常数量大，才能快速标注数据、训练出更好的算法模型，采用“非正则”模式，否则仍需要继续使用“关键词+正则”模式。

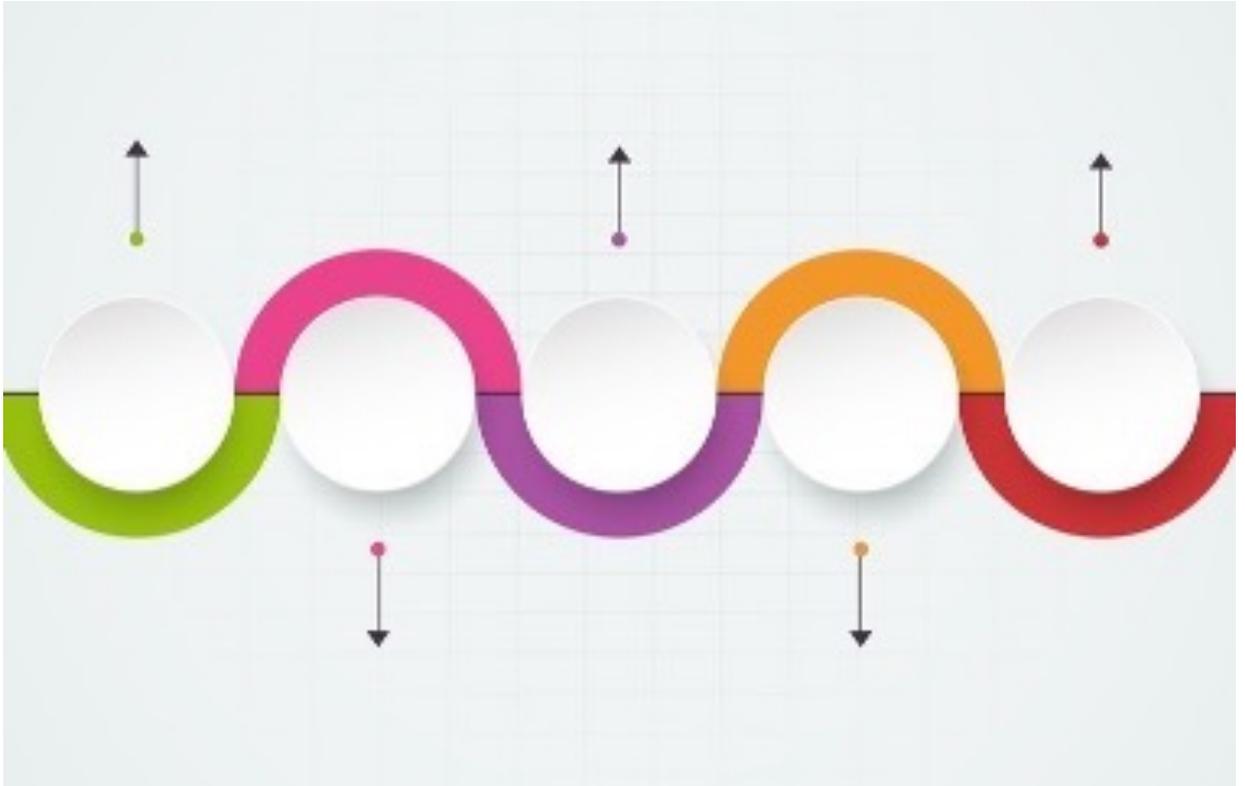
其次也要考虑到质检项的成熟度——太新的、标准尚不明确的质检项，适合先用“关键词+正则”模式迭代起来，再决定何时升级到“非正则”模式。综合来看，在实践中采用“双模”方案，让两种模式各司其职，可最大程度提升质检效率。

双模质检，各司其职。“非正则”质检方式通过引入深度学习算法模型，能够更好地利用上下文语义来判断一个句子是否命中了质检项，显著提升找全率（术语“召回率”）和找准率（术语“准确率”）。

同时，由于深度学习算法模型对训练数据量有要求，因此“非正则”方式更适合目标通话量较多的质检项，而目标通话量过少的质检项，无法训练出好的算法模型，需要继续使用“关键词+正则”的方式。



综合来看，在实践中采用“双模”方案，让两种模式各司其职，可最大程度提升质检效率。



### 三、新实践：SOP流程质检

如何监督业务人员在遇到特定场景或客户的特定问题时，有没有执行标准流程，有没有按要求进行回复？

## 什么是“流程质检”？

流程质检的主要目标是，判断员工与客户沟通时“是否按照规范流程，回答了客户的问题”。

具体来看，流程质检的质检项，首先需要判断一个对话是在讲什么事情，然后再判断针对这件事情，业务人员的回答是否满足一个或多个规范条件。

其实，也可以认为“流程质检”是一种 2 个以上质检项（语义点）的复合质检。复合质检若想做到比

较好的效果，需要每个语义点都能实现较高的召回率和准确率。

因此，在传统的“关键词+正则”模式之外，引入效果更好的、基于深度学习模型的“非正则”模式来生产质检项（语义点）势在必行。

下面，我们来看两个实际的例子，帮助大家理解循环智能的“流程质检”功能是如何工作的。

**当** 会话命中  时

**则** 须同时满足以下语义点条件

1 **出现**  X

2 **出现**  X

**+** 添加语义点条件

## 实例：“单条件”流程质检

“单条件”流程质检指的是，当客户提到某个具体的场景，业务人员在回复中必须出现某个语义点（单个）。例如，很多企业的服务规范，要求业务人员遇到“客户要求反馈”的场景时，及时作出“安抚解释”。如果没有出现“安抚解释”的语义点，即判定该客服或销售人员未按流程提供服务，须减去一定的分值。

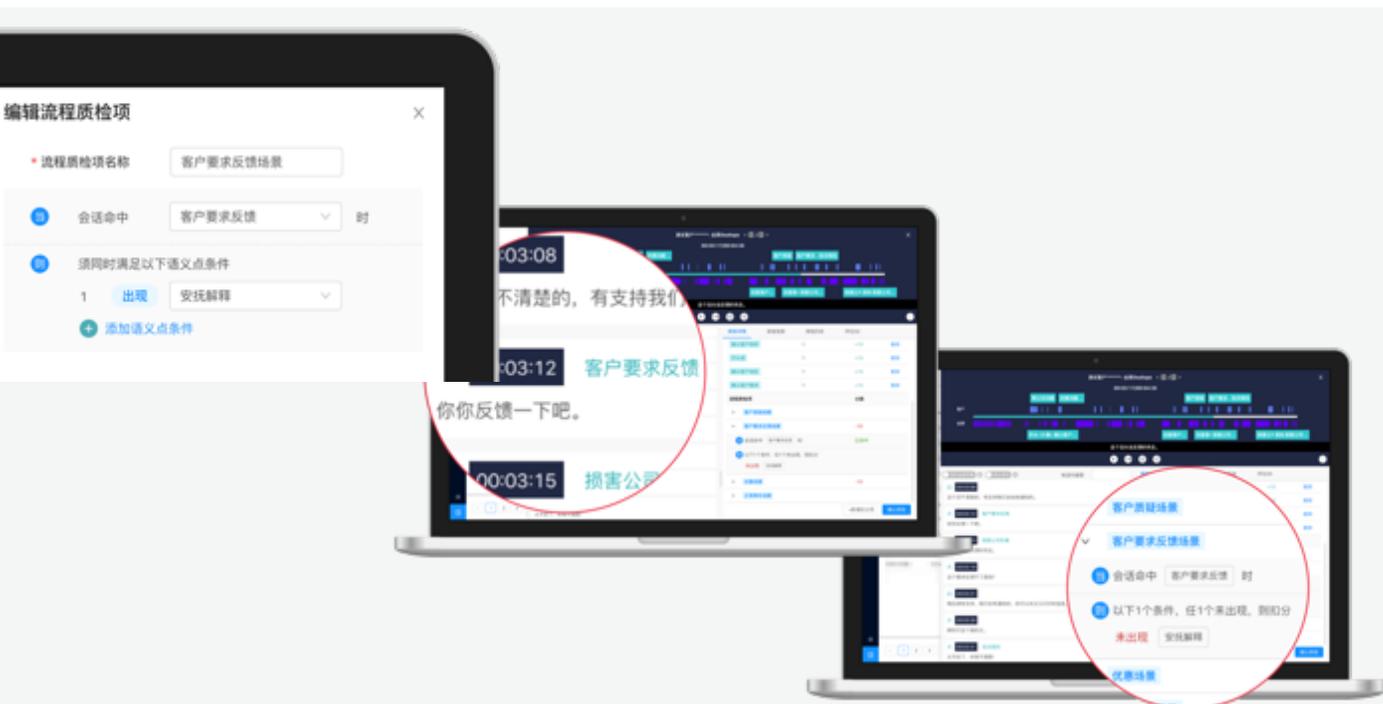
首先，我们要配置好“客户要求

反馈场景”的流程质检项。

配置完成之后，在机器质检系统中搜索“客户要求反馈场景”，就可以找出所有命中该场景的不合格对话。

在这个例子中，机器质检系统判定在这通对话中，客服人员没有进行“安抚解释”，所以要扣除一定的分数。

最后一步，人工进行复检，看看机器有没有误判或漏判。



## 实例：“多条件”流程质检



“多条件”流程质检，指的是当客户提到某个具体的场景，业务人员在回复中必须出现多个特定的语义点。例如，某个专业的、提供居住服务的企业，要求当客户提到“空调不运转”的场景时，在线客服人员需要向客户确认“插座是否通电”、“物业地址”和“维修时间”，如果没有提到其中任何一个，都会被判定为不合格。

同样，首先我们要配置好“空调

不运转场景”的流程质检项：

配置完成之后，在机器质检系统中搜索“空调不运转场景”，即可找出所有命中该场景的不合格对话：

在这个例子中，在线客服人员没有提到“插座是否通电”，因此被判定为不合格，需要扣除一定的分数：

最后一步，同样是人工进行复检，看看机器有没有误判或漏判。

## 监督不是目的

想了解一家企业是否值得信赖，打一次热线电话就知道了。

电话那头的客服或销售人员，代表企业直接与客户打交道，他们能否快速定位和解答问题，往往可以反应出企业的专业性和规范性。为了赢得更多客户信赖，企业不断总结出客户常见问题的标准作业流程（SOP），要求业务人员按流程帮助客户定位和解答问题。

越重视客户体验的中大型企业，往往对客服人员的规范执行情况要

求越严格，对“流程质检”的需求比较大。最近，我们发现，有越来越多企业也开始重视销售场景的标准作业流程（SOP），因此，通过“流程质检”功能监督销售人员执行力的需求，也逐渐增多了。

不过，无论是面向客服人员，还是销售人员，“监督”从来不是目的。对企业而言，找到更好的方法监督员工的执行力，最终还是为了提升员工的执行力和企业的运营效率。



*“对企业而言，找到更好的方法监督员工的执行力，最终还是为了提升员工的执行力和企业的运营效率。”*



## 四、质检项优化

为了使机器质检和人工复检协同工作，提升整体质检工作的效率，我们需要对不同的质检项，采取不同的优化策略。

## 召回率和准确率的关系

我们希望每个质检项所命中的目标通话又全又准。这也是循环智能的质检产品引入“非正则”模式的原因。

但在优化质检项的命中效果时，当召回率和准确率都达到了一定的高度，想要继续提升，难免会遇到“二选一”的问题：召回率和准确率互相影响、此消彼长，一个指标增长，另一个指标通常就会下降。

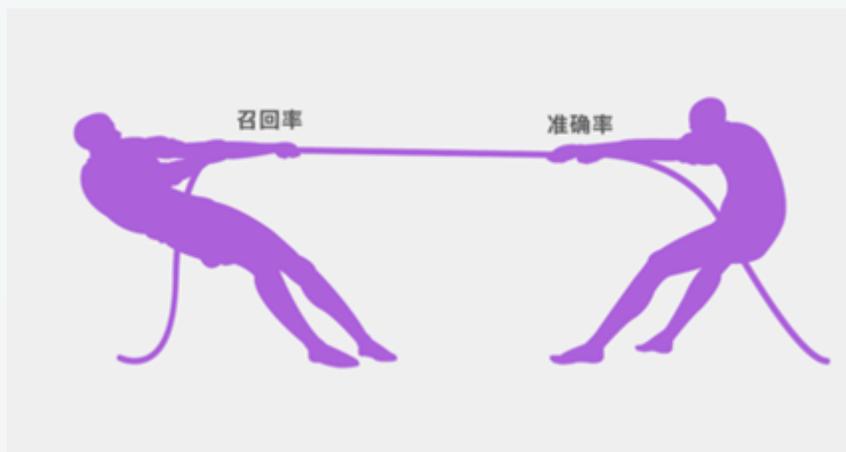
在实际应用中，大部分企业的业务流程是“机器质检+人工复检”：将机器质检的结果交给人工做复检。因此，机器质检的召回率和准确率

变化，会影响到人工复检的成本：

- 提升召回率（适当牺牲准确率），意味着降低“漏检”的风险，但可能会增加“误检”带来的人工复检成本；

- 提升准确率（适当牺牲召回率），意味着降低“误检”带来的人工复检成本，但可能会增加“漏检”的风险。

那么对于不同的质检项而言，如果碰到“二选一”的问题，该如何做出选择？



## 负向质检项：通常“召回率”优先

在销售、客服质检中，负向质检，通俗地讲，是指找出业务员“说了什么不该说的内容”。负向质检是企业比较常见的需求，尤其是在受监管比较严格的领域。

在贷后资产管理（催收）领域，质检项以负向为主，例如恐吓威胁、疑似私收钱款等。

负向质检项通常有两个特征，第一是违规量通常都不太大（大部

分通话不会违规），第二是漏检之后的风险相对比较大。

因此，针对负向质检项，我们通常应该通过调整算法模型的参数或规则代码，保证“召回率”优先，将更多涉嫌违规的通话都找出来，然后增加人工成本去做复检，“宁可错杀一片，不可放过一个”。

宁可错杀一片，  
不可放过一个

## 正向质检项：通常“准确率”优先

在销售、客服质检中，正向质检，通俗地讲，是指找出业务员“该说的什么内容没有说”。

企业可以选择不同的管理理念，可以是“惩恶”理念，对业务员做得不好的地方进行减分；也可以是“扬善”理念，对业务员做得好的地方进行加分。最近几年，“扬善”理念越来越受到企业重视，这种方式可以用来激励业务员变得更加专业、更加规范，更有利于形成正向循环。

例如客服领域的规范性用语：

“标准开场白”、“标准结束语”、“服务延伸用户（请问还有什么可以帮您）”以及“确认客户预留信息”等。

这类正向质检项，通常目标通话量比较大，如果错误率较高，复检成本就会很高。

面临二选一的时候，我们通常应该通过调整模型参数或规则代码，优先提升准确率，降低人工复检成本，“宁可漏掉几个，不可错杀一片”。

**宁可漏掉几个，  
不可错杀一片**

## 实际场景更复杂

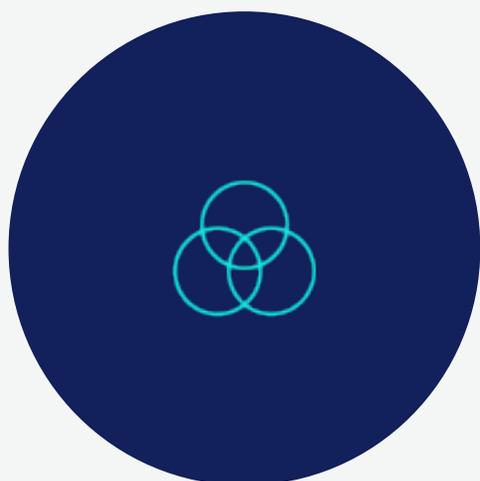
需要重申，我们在优化质检项命中效果的时候，最好的情况是，能找到同时提升召回率和准确率的方法，或者能找到大幅提升其中一个指标，而另一个指标不会明显下降的方法。如果这两条路都走不通，为了继续提升效果，才需要结合漏检的风险高低以及人工复检的成本，进行“二选一”。

通常，因为负向质检项的漏检风险比较高，所以“召回率优先”，降低漏检风险；而正向质检项的目标通话量比较大，所以“准确率优

先”，降低复检成本。

不过，也有少量例外。比如有的负向质检项，目标通话量不算少，而且风险也比较低，那可以考虑“准确率优先”；有的正向质检项，目标通话率非常少，而且重要性比较高，那就应该考虑“召回率优先”。

此外，在实际应用中，还要将其他因素综合考虑进去，比如不同质检项的召回率和准确率提升难度不同，比如不同企业的复检员数量是不同的……我们才能更好地制定不同质检项的优化策略。



*“通常，因为负向质检项的漏检风险比较高，所以召回率优先，降低漏检风险；而正向质检项的目标通话量比较大，所以准确率优先，降低复检成本。”*

## 数字背后是“人心”

相比“关键词+正则”的方案，企业引入“非正则”质检模式，意味着可以发现更多问题。这是新一代深度学习技术的优势，算是一种技术红利。

我们的客户成功团队，每个月向客户做汇报的时候，也会着重展示这方面的效果。不过，在服务了几个月之后，我们却发现“效果”不像刚开始那么好了，之前能多找出几倍的问题，后来这个数据逐渐不那么显著了。按道理，我们使用了自动训练的功能，企业人工复检的结果会反哺给算法模型，让算法模型越用越准，不应该出现效果更差的情况。

直到我们仔细审查了数据，才发现每个质检项的效果都在变好，找准率和找全率都在上升，但找到的问题确实更少了。真正的原因其实是问题本身真的更少了，“巧妇难为无米之炊”。而之所以问题会变少，

是因为发现了更多的问题，而且企业也能对应找到更多解决问题的方式。

我们换到业务员的视角，更容易理解其中的逻辑：业务员很快就发现，新的质检系统很容易揪出自己犯错的地方，所以自然会更加谨慎，减少犯错，以防被罚款（很多企业采用了惩罚）。也就是说，新的系统配合罚款这种行之有效的管理方式，对业务员产生了威慑力，导致犯错减少。

**从发现的问题越来越多，到发现的问题越来越少，数字的变化，背后其实是“人心”的变化。**

我们观察发现，每家企业的“拐点”出现的早晚都不太一样，主要取决于企业业务稳定性（新的质检项需要迭代出更好的结果）和人员的稳定性（新人更容易犯错）。在比较理想的情况下，“拐点”通常会在第六个月前后到来。



## 五、微信质检

企业与客户沟通的渠道处于不断变革中。当前，企业微信被越来越多企业采用。因此，我们在电话录音质检和在线IM文本质检的基础上，新增了对企业微信的支持。

## 企业微信需开通“会话存档”

当前，企业微信被越来越多企业采用，员工采用企业微信可与客户的个人微信进行消息互通，同时可展示企业认证员工的身份，有助于赢得客户的信任。

例如：银行的理财顾问使用企业微信，向高净值客户推荐理财产品；婚恋平台的红娘使用企业微信，为会员提供一对一服务；高端品牌商的导购员使用企业微信，向顾客推荐新到的商品。

对企业而言，员工通过任何渠道与客户进行的沟通，都应该保证沟通的质量和规范性。因此，我们在电话录音质检和在线IM文本质检的基础上，新增了对企业微信的支持。

使用企业微信的客户，为员工开通“会话存档”功能之后，即可调用官方接口，将会话内容接入循环智能的质检产品中，对沟通中的文本和语

音（循环智能会将语音转写为文本）进行质检。需要注意的是，就像电话录音需要提醒客户一样，开通了企业微信的“会话存档”功能之后，企业员工在与客户沟通时会自动出现“服务须知”的提示。

通常语境中的质检，是指检查整个对话中是否出现违规内容，例如在整个对话中的任何时候出现“辱骂或威胁客户”都是违规的，即“说了什么不该说的”。

不过，现在企业的流程规范化程度越来越高，也需要考核员工是否遵循了SOP流程规范，即“该说的有没有说”以及“说的时机是否正确”。

例如，当客户提到A情况时，员工是否按照规范回复了B内容；或者当员工提到C内容之前，是否确保客户已确认了D信息……

## 微信质检的常见问题FAQ

# Frequently Asked Questions



**问：微信聊天中的图片和表情可以质检吗？**

答：目前仅针对文本和语音，图片和表情暂不处理。

**问：支持群聊、企业微信与企业微信之间的会话质检吗？**

答：企业微信暂不支持上述会话内容的存档。

**问：跨越多天的会话内容，可**

**以每天只质检增量部分吗？**

答：循环智能的系统支持仅质检每日增量部分，也支持将多天内容整合在一起进行质检。

**问：语音是微信转写的，还是循环智能转写的？**

答：循环智能提供短语音转写为文本的服务，可支持对专有领域内容做算法优化。



## 六、保险行业的实践

在服务保险客户的过程中，循环智能逐渐形成了保险质检的“三级火箭”解决方案。

## 传统机器质检效果不佳

为了满足政府监管和企业自身提升服务质量的需求，过去几年，各大保险公司、保险中介机构纷纷部署了机器质检系统、双录质检系统，以期能够及时发现和纠正，销售员与客户沟通过程的不规范之处，降低企业被投诉带来的合规风险。

但保司和保险中介机构渐渐发现，过去部署的机器质检系统效果无法令人满意。其中关键原因有二：

第一，人与人沟通时讲的话是高度个性化的、表达方式是复杂多样，仅仅依靠“关键词和正则”进行穷举的传统机器质检方案，很难识别出真正的句子语义。例如，在质检项“夸大陈述”中，保司发现销售员夸大重疾险保障范围的方式是穷举不完的；在质检项“不当对比”中，保司发现销售员通过将保险与其他同业产品对比，以及与储蓄、基金、股票等金融产品对比来强调产品优

势情况也非常多，用规则穷举的效果很差。

第二，保险销售的业务流程、逻辑是复杂的，对销售员的要求非常精细化，传统的机器质检方案，无法对包含复杂场景和精细化要求的质检项进行监督。例如，在质检项“意外医疗告知不严谨”中，需要先筛选出涉及到意外医疗的会话，再监督销售员是否明确告知是意外导致，以及是否明确告知意外医疗适用的保障范围。传统机器质检很难处理这种需要对业务流程规范进行监督的情况。

为了破解这两大难题，一些领先的保司和保险中介机构，率先落地了新的解决方案，通过将基于AI技术的新一代合规质检产品部署到业务流程中，大大提升了质检的效果和效率。具体来看，他们采用了三种新的实践。

## 实践一：从“关键词”升级到“一段话的语义”

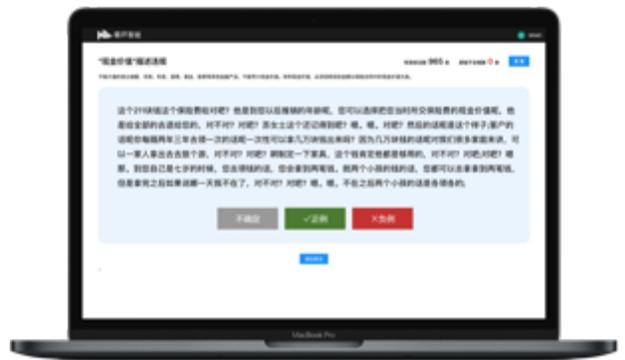
从工作原理上看，保司和保险中介机构部署的初代机器质检系统，通过将录音和微信语音转写成文本，然后借助“关键词和正则表达式”进行穷举，以查找其中可能涉及违规的会话。

而新一代的AI机器质检方案则与之不同。新方案通过“喂”给机器足够多的违规实例片段和不违规实例片段（即经人工判断属于某项违规或不属于某项违规的对话语句片段，这个过程在AI领域被称为“人工标注”），训练机器算法去“学习”违规对话片段的隐含特征，然后用“训练”出的模型识别更多对话片段的语义，看看是否命中这个“AI语义点”。

质检项可能包含一个“AI语义点”，也可能包含多个“AI语义点”的组合。保险质检项的特点是专业程度很高，例如质检项“现金价值描述违规”由单个AI语义点组成，但其定义需要高度依赖保险专业知识和行业经验，只有仔细检查了超长的上

下文对话之后，才能准确得知一段对话是否命中了“现金价值描述违规”。

这也是循环智能（Recurrent AI）在服务多家头部保司的质检项目时，学到的第一条经验：在“质检项”生产过程中的人工标注环节，需要保险企业的人员直接参与进来，与AI质检供应商内部的保险行业专家合作进行生产（标注数据），才能更快提升模型的准确性。第二条重要的经验是，保司和提供新一代AI合规质检解决方案的公司，要花很多精力来对齐质检项的定义，这部分的工作越完善，后续的工作效率就越高。



## 实践二：引入自定义字段和复杂流程&逻辑配置

保险销售过程的合规质检，至少包含了两种基本的情况：

第一种，发现销售员“说错话”的情况，例如“夸大陈述”这样的质检项，通常只包含单个AI语义标签（AI语义点、AI语义画像），只要说了就算违规；

第二种，发现销售员不严谨或不规范的地方，例如“产品介绍遗漏”这样的质检项，通常会涉及多个标签（AI语义点或正则语义点），因为先要判断当前对话的场景是否涉及某个长期险或短期险，然后再判断是否讲到了相应保险产品的所有五项或七项特点。

事实上，第二种情况还可能涉及到多种复杂场景，有时候需要引入企业的自定义字段，有时候需要判断对话内容上下文的逻辑，才能判断对话是否存在不严谨或不规范的地方。例如：

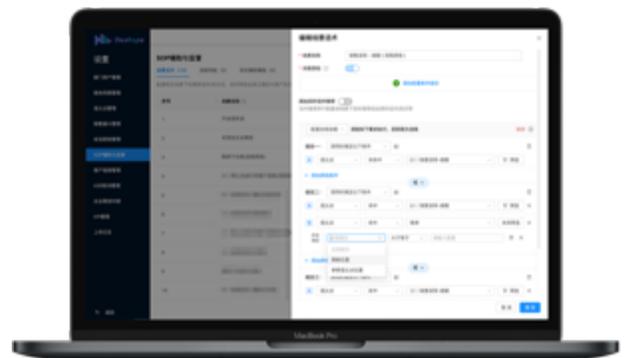
1) 不同产品的“责任免除”范围是不同的：因此需要引入企业的自定义

字段，先判断这通对话是在介绍哪个产品，然后再判断相应的“责任免除”范围是否准确陈述。

2) 一些违规项是由多个语义点的逻辑关系组成的：例如，保单递送场景，提及电子保单或纸质保单其中一个都算合规。但是关于“保单生效日”陈述的要求是，必须讲到生效日，同时提到“扣款不成功不生效”，才符合要求。

3) 先后顺序：在投保确认前，不允许核对客户的个人信息（信用卡号等）。

循环智能的新一代合规质检系统，支持企业自定义字段，以及复杂多样的质检项场景逻辑配置，对于复杂逻辑质检项的支持更加精细化、更加完善。



## 实践三：引入秘密武器“违规值”评分

在服务保险企业的过程中，循环智能（Recurrent AI）团队发现，尽管使用了AI语义点，也使用了多个AI语义点的逻辑关系配置功能，但有一些质检项过于复杂（例如：混淆新老保单），AI语义点加上逻辑关系组合也难一网打尽。

面对这种棘手的情况，循环智能（Recurrent AI）创造了一种新的“违规值”质检方式。这种新方法，并不关心具体哪句话违规了，而是专注计算整个会话违规的可能性。

通过把整个对话（甚至同一个保单对应的多个对话）当成一个整体进行机器学习：在训练阶段，将“违规”的整体对话和“没违规”的整体对话输入算法模型，学习“违规”

对话的隐藏特征；在执行阶段，将新的对话输入算法模型，然后计算该对话违规的概率。根据企业的实际需求，违规值计算又包括两种场景：

- 整个会话命中单个质检项的违规值
- 整个会话命中多个质检项的违规值（例如命中S级质检项组合中的任意一个算违规）

在需要100%人工复检的保险成功单质检场景，如果在算法建模阶段经过了充足的数据训练，那么循环智能（Recurrent AI）的“违规值”排序功能，基本上可以在违规值前60%的会话中找到大部分违规对话，最高节省40%的成功单质检工作量。



# 保险质检的“三级火箭”

在服务保险客户的过程中，循环智能 (Recurrent AI) 逐渐形成了保险质检的“三级火箭”解决方案：

- 第一级：从字和词级别的关键词+正则方式，到引入句子和段落级别的AI语义点方式；
- 第二级：引入企业自定义字段和逻辑规则判断，实现多个语义

点之间复杂的流程&逻辑判断；

- 第三级：计算整通会话违规的可能性高低，通过违规值排序，帮助人工复检员节省工作量。

通过引入和部署经过考验的AI质检技术，保司在质检工作中的效率可以得到数倍提升。





←  
扫码预约  
产品演示

 [www.rcrai.com](http://www.rcrai.com)

 010-56271640

 [market@rcrai.com](mailto:market@rcrai.com)

循环智能 (Recurrent AI) 是一家专注于销售科技SalesTech领域的企业服务公司，致力于让企业与客户的每一次沟通有更好的成效。循环智能创立于2016年，采用北京和上海双总部的模式，并且在深圳、无锡设有分公司。



北京



上海



深圳



无锡

循环智能的创始团队来自清华大学、卡内基梅隆大学，并获得博裕投资、红杉中国、真格基金、金沙江创投、靖亚资本、万物资本和华山资本等知名投资机构的支持。













