

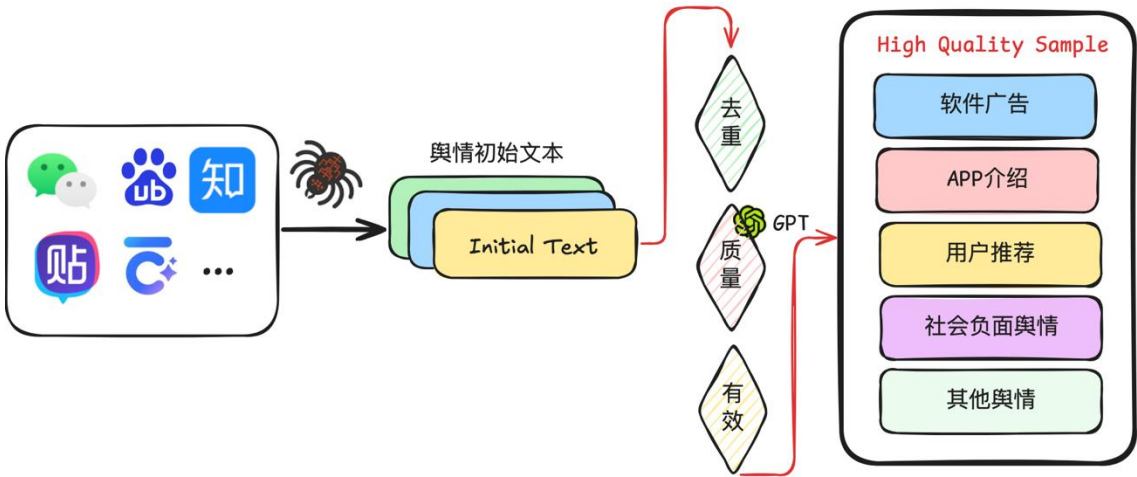
# 无监督新型非法集资自动标注框架

**背景：**目前互联网上存在**大量**具有非集风险的**舆情文本**(未标注)，同时专家组内部具有一套**完整且复杂的****标签体系**。但是如何将每条舆情文本分类到具体的标签中，成为了亟需解决的问题。

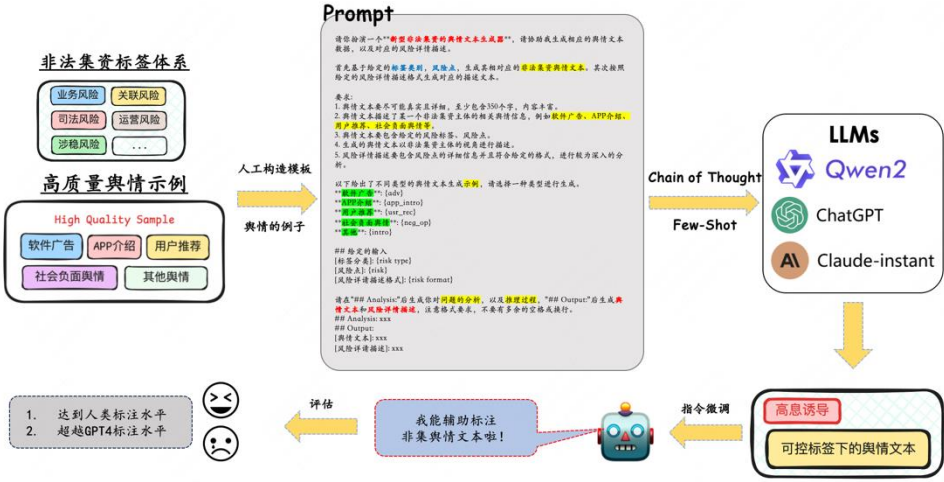
## 挑战：

- 1. 目前人工标注存在大量人力成本，一批高质量的标注人员需要能够理解复杂的标签体系以及标注日益增加的舆情文本。
- 2. 互联网上的舆情文本具有质量低、语义重复并且充斥大量无关信息等特点。同时随着大模型的发展，日后的互联网上的舆情文本会倾向于利用LLM生成的更高质量文本。

**方案：**因此，我们提出了一个基于大模型的无监督自动标注框架。利用**大模型蒸馏技术**，**生成**可控标签下的高质量舆情文本，再根据**(高质量舆情文本text, 可控标签y)**训练一个**辅助**标注模型m，最后在初始舆情文本数据集上进行标注。



数据预处理



自动标注框架

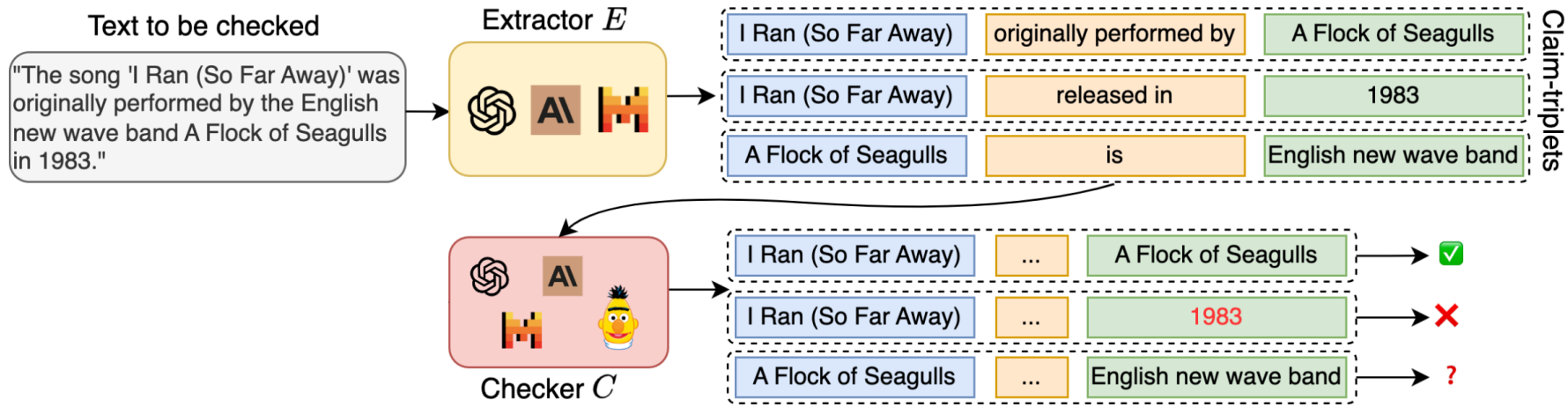
# 细颗粒度非法集资风险预警模型

**缺陷：**自动标注框架中的合成数据目前是基于**单一可控标签**生成的舆情文本，因此训练出的标注模型**只能生成单标签**，但是在实际场景中，一段舆情文本会包含**多种**风险标签。因此我们提出了一个**细颗粒度的预警模型**。

**目标：**利用一个**打分模型**，对给定的舆情文本进行非法集资风险的打分，检测互联网上的文本是否有潜在的风险，识别**风险具体的类型**以及**相应的风险等级**。

**方案：**

- 1. 首先利用大模型Extractor E对舆情文本进行摘要分割，分为多个三元组Claim。
- 2. 训练一个标注模型Checker C，对每个三元组进行风险类型分类，判断其a.是否具有风险（3分类 风险 | 无风险 | 不确定） b.具有哪一类风险（风险标签体系）。
- 3. 最终舆情文本的风险等级=被标注为风险claim的条数，风险具体的类型=[claim1: risk\_type1, claim2: risk\_type2, ...]



谢谢！