

# 实习工作总结



AI-Lab-NLP 内容质量项目  
-- 何逸宸

- 伪科学内容识别/健康专审
- 实体链接
- 信源库构建
- 通用多模态审核模型

# 内容质量项目

## 文本二分类

# 项目背景和原因

虚假信息的发布和传播会对社会造成很大的危害

- 影响对重大事件的认知

## 俄乌冲突

- 俄罗斯国防部称，波兰军队已经进入到了乌克兰，并且正式加入了对俄罗斯的战争，俄罗斯军队表示，不会让波兰军队走掉，如果有必要的话，俄罗斯军队将立即对波兰发动闪电攻击·最新消息
- 12000美军进入，炸毁俄罗斯石油管道！普京下令紧急出兵....
- 俄军开火立陶宛，48枚导弹凌空爆炸！这就是惹怒普京的代价
- 威胁健康，造成财产损失



# 项目方法

风格检测	谣言检测	事实检测	用户评论/行为	传播模式挖掘
<ul style="list-style-type: none"><li>简单，搭建模型成本较低</li><li>召回率低</li></ul>	<ul style="list-style-type: none"><li>构建谣言库，可抽象为自然语言推断</li><li>手工录入谣言，无法发现新谣言</li></ul>	<ul style="list-style-type: none"><li>可发现新的不实信息</li><li>技术挑战大</li></ul>	<ul style="list-style-type: none"><li>特征容易挖掘</li><li>不实信息已经对用户造成了影响</li></ul>	<ul style="list-style-type: none"><li>获取标注数据成本较低</li><li>检测准确率依赖于传播规模，不能及时发现谣言</li></ul>

# 伪科学内容识别

# 伪科学内容识别 文本二分类

# 保持科学理性的思考方式

## 1. 伪科学谣言为什么会被传播

在科学技术这个大的领域内，虚假问题很容易泛滥成灾。但由于披着“科学”外衣，往往具有更强的迷惑性，加之很多人科学素养不足，辨别能力有限，出于“宁可信其有，不可信其无”的避害思维，甚少刨根问底、追问究竟，所以一看是科学相关的“高级”内容，就失去了自己的判断能力，被牵着鼻子走，受了骗还不知道。

同时，科学类谣言往往伴随着一定的危险性，加上公众的恐慌心理、从众心理和心理定势，容易产生伪科学类谣言的传播温床，最后再通过网络世界强大的聚合、放大效应，容易迅速爆炸性传播、病毒式感染，最后产生大范围大批量的伪科学“信徒”。比如这些谣言有不少的拥护者：地球是平的，百慕大三角一直发生神秘事件，地球深处存在地心人。

## 2. 识别科学谣言对个人的意义

聚焦到科学科技的层面上，准确有效地判断其中的虚假内容，对我们也有着多重意义：从个人方面说，可以让我们快速地判断一个内容的真假，从而在信息洪流中保持自己不被带偏不受蒙骗，同时不做谣言的传播者，进而做出一些理智的选择。

通常情况下科学类的谣言，都包装的比较高端，也会和现在的互联网行业、前沿科技挂上一些边。镰刀从来都是挥向肥的韭菜。能够识这类虚假谣言的话，就可以少交一些智商税，不会因为什么量子治疗、WIFI致癌、防辐射服这种生活中的谣言所欺骗，从而造成一些利益上损失。

此外，分辨出虚假内容的思路方法，其实也是科学思维建立的一种搭建过程，可以帮助我们在信息洪流中，面临多个分岔路口时，能保持自己的科学思考和理性判断，从而做出理智又正确的抉择，不至于慌乱到无所适从。

## 3. 识别伪科学对工作的帮助

在另一个维度上，对科学类虚假内容的判断，对我们的工作内容也会带来一些收益。如果你立志成为这方面的研判专家，快速准确的判断是一项基本功。如果你是其他业务线的同学，快速地定位和判断出科学虚假问题，会帮助你在做决策的时候考虑到更多的风险，打开更大的视野。

古语说，知己知彼百战百胜。掌握了对科学虚假内容的判断方法，可以再面临此类问题时，制定相对应的对策和方法，对有害内容进行阻击，可扩大优质内容的运营、传播、推广等动作，同时在谣言治理，危险账号拦截等方面有一些针对性的对策。

# 伪科学内容识别 文本二分类

## 1. 伪科学谣言为什么会被传播

“去氘水修复DNA”

清清宇泉  
2022-5-11 15:18 · 来自广东

当人体新陈代谢功能受损时，适当饮用去氘水，可对DNA有很好修复的作用。去氘水——是运用现代科学技术，把自然界中的水去除部分氘后得到的水，也叫贫氘水，低氘水，超轻水。科学研究发现，饮用水的氘含量越低，水的活性越强，容易进入细胞，被生物体吸收利用，对保持身体健康具有重要意义。



1336展现

评论 1 赞 | 0 转发

暂无评论，点击抢沙发

写评论...



“院士只会讲脱口秀”

彗星撞火星  
2022-1-2 08:34

霍金早就警告过人类不要接触外星人，但某位国内的学者硬要公开演讲说要试图联系外星人？

这位国内学者兴奋说道国外最近探测到了外星人的信号源，并且表示要快点建立连接。。。 (无语)

怪不得出不了一个诺贝尔，别人的科学院士拿诺贝尔，国内的院士讲脱口秀，还不是自己吹嘘并研究的。也不管霍金说过什么！

这难道就是差距吗？#我要上头条#

5748展现

评论 2 赞 | 0 转发

彗星撞火星 1 赞

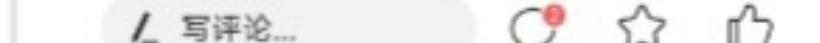
但凡能来地球我们就直接投降好了

回复 > 1-2

彗星撞火星 1 赞

差距太大

回复 > 1-2



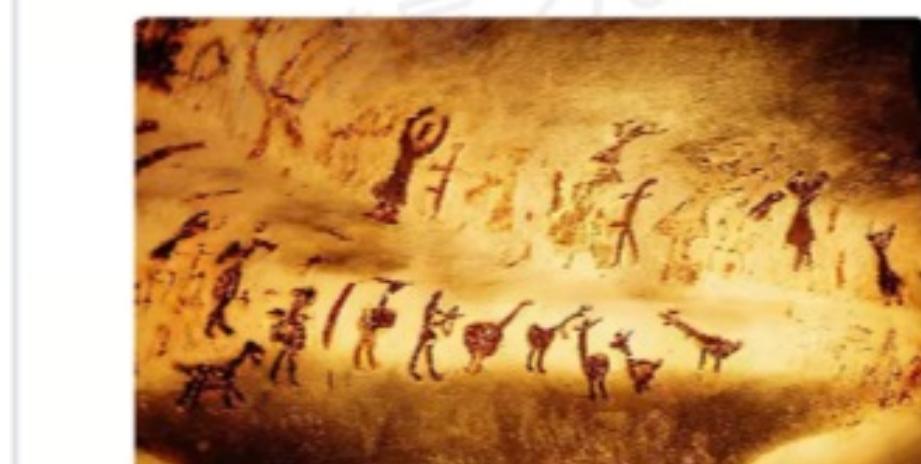
“外星人到过地球”

外星人在5000年前就出现过？史前壁画记载外星人早已拜访地球

岛主阿坤 1 赞

2021-11-16 12:13

这一切神秘事件有些与牛鬼蛇神有关，有些与物理现象有关，有些与外星人有关。今天我们要讲的，是与外星人有牵连的事情。大多数情况下，将某件事视为外星人访问地球只是人类的一种幻想，而目前人们所能做的只有确认这神秘本身的真实性。



今年四月，美国方面已经证实，在14到15年间，有许多UFO在其军事基地上空被发现。对

于这些是否为外星人，美国军方表示暂时无法证实！然而，从目前的数据来看，地球上并没有外星人，或者可以说外星人从未访问过地

球。



# 保持科学理性的思考方式

“高速物质形成金属氢”

金童希瑞  
2021-12-27 15:33

小行星俯冲瞬间高速流动的物质转化为金属氢，金属氢聚合形成的二氧化硅以及二氧化硅衍生的硅酸盐会记录小行星俯冲瞬间形成的磁场。

可见，用古地磁确定地质构造运动是极其荒唐；而根据古地磁得出地轴翻转更是无稽之谈！



748展现

评论 6 赞 | 0 转发

暂无评论，点击抢沙发



“地球在不停地摇摆”

摇摆的地球  
2021-12-27 06:22

地轴倾斜角是地球摇摆运动形成的，倾斜角度即是地球的摇摆幅度，地轴以约183天为一个周期缓慢旋转，致使地球出现四季变化。伴随地球公转，地球要完成两个摇摆运动周期，地球表面才能出现一次四季变化，如果没有公转，只有自转和摇摆运动，只需完成一个摇摆运动周期，地球表面就能出现一次四季变化。

四季变化是地球摇摆运动形成的，地轴倾斜方向是不停的匀速的旋转改变的，所谓的“地球公转形成四季变化”理论和“地轴总是倾斜的指向同一方向”理论都是绝对错误的。

地球摇摆运动关系人类生存，根据地球摇摆运动形成原理可知：两极冰川全部融化，地球质心位置会发生转移，地轴位置重新确立，地球自转发生翻转，导致第六次生物灭绝事件发生。根据现在气候恶化速度，两极冰川全部融化指日可待，人类生存问题迫在眉睫。有效控制两极冰川融化，阻止地球质心位置转移，是人类继续生存唯一出路，科学合理治水是唯一选择。如果还抱着温室效应这一错误理论做事，人类将走向灭亡，请全世界科学家们重视！

4866展现

评论 2 赞 | 1 转发

摇摆的地球 1 赞

上一轮气候周期变化即将结束，下一轮气候

写评论...



# 伪科学内容识别 文本二分类

## 写作风格

教育教学

写真地理

2020年6月 第22期

### 熟鸡蛋变成生鸡蛋(鸡蛋返生)——孵化雏鸡的实验报告

郭 平 白卫云

(郑州市春霖职业培训学校 河南 郑州 450000)

**摘要:**“鸡蛋返生”，顾名思义，就是由熟鸡蛋再变成生鸡蛋。这是一个难以想象的，甚至是不可见的，但是这样奇特的现象确实发生在郑州春霖职业培训学校发生了。一群特别培训的学生，在郭平老师的指导下，正在进行一个奇特实验，即熟鸡蛋重新变成生鸡蛋，并将返生后的生鸡蛋进行孵化或雌雄。并且已经成功返生了40多枚。

**关键词:**生鸡蛋；熟鸡蛋；鸡蛋返生；孵化雏鸡；实验报告

【中图分类号】S831

【文献标识码】A

鸡蛋奇特返生的现象，根据鸡蛋的组织结构及功能。鸡蛋经过高温100℃开水煮20分钟，变成熟鸡蛋后，学生们运用自己的超心理意识能量方法等，将这些熟鸡蛋变成生鸡蛋，现在我们将这种奇特现象分享给科学探秘爱好者，共同探究其内在理论依据。

实验材料：鸡蛋10枚，一次性纸质茶杯10个。

实验场地：郑州春霖学校507教室。

实验时间：2020年6月12日11时。

室内温度：摄氏25℃，保持室内安静。

参加人员：郑州春霖学校特训生10人，见表1。  
表1 观察见证专家及学生家长

姓名	职务(单位)	备注
郭平	校长	郑州市春霖职业培训学校
白大勇	教授	香港毕业大学书记
高建民	院长	郑州大学设计院系分院院长
单松柏	原国家地震局郑州物探中心主任	地震学家
白玉忠	主任医师	河南中医药高等专科学校附属医院
尹杰		山西中医研究所
高爱华	家长	春霖职业培训学校
陈静霞	家长	孙家庄小学
赵真真	家长	碧水蓝天小学
郭大勇	教授	新郑市第二实验小学



钙等矿物质组成。对鸡蛋内容物起保护作用。①外层卵壳膜：是层无结构纤维膜，主要是保护鸡蛋内容物水分不丢失。②内层卵壳膜：是一种可透气膜，空气可以进出。③气室：位于鸡蛋的钝端内（在大头与卵壳膜之间），是由内层卵壳膜与外层卵壳膜之间窄分开形成一个小气室，贮存空气，具有胚胎发育时供应其呼吸功能<sup>[1]</sup>。④系带：卵黄的两端由浓稠的蛋白质组成卵黄系带，其功能是维持卵细胞共定于蛋白中心位置，对卵细胞具有起着缓冲作用，可防止卵细胞的震荡。⑤卵黄膜：位于卵白与卵黄之间的一层薄膜，是卵细胞的组成部分。⑥卵黄：呈黄色，位于细胞的中央，是鸡卵胚胎发育的主要营养物质。⑦胚盘：位于卵黄表面中央有一圆形盘状小白点，呈椭球状等。⑧卵白：卵壳膜与卵黄膜之间，为胚胎发育提供水和营养物质。卵黄表面中央有一圆盘状的小白点（就是在蛋黄上看到的小白点）称为胚盘，里面含有细胞核，未受精的卵，胚盘色淡而小，已受精的卵，胚盘色浓而略大，这是因为胚胎发育已已经开始<sup>[2]</sup>。如果又是受精卵，胚盘在适宜的条件下就能孵化出雏鸡，胚盘进行胚胎发育的部位<sup>[3-5]</sup>。

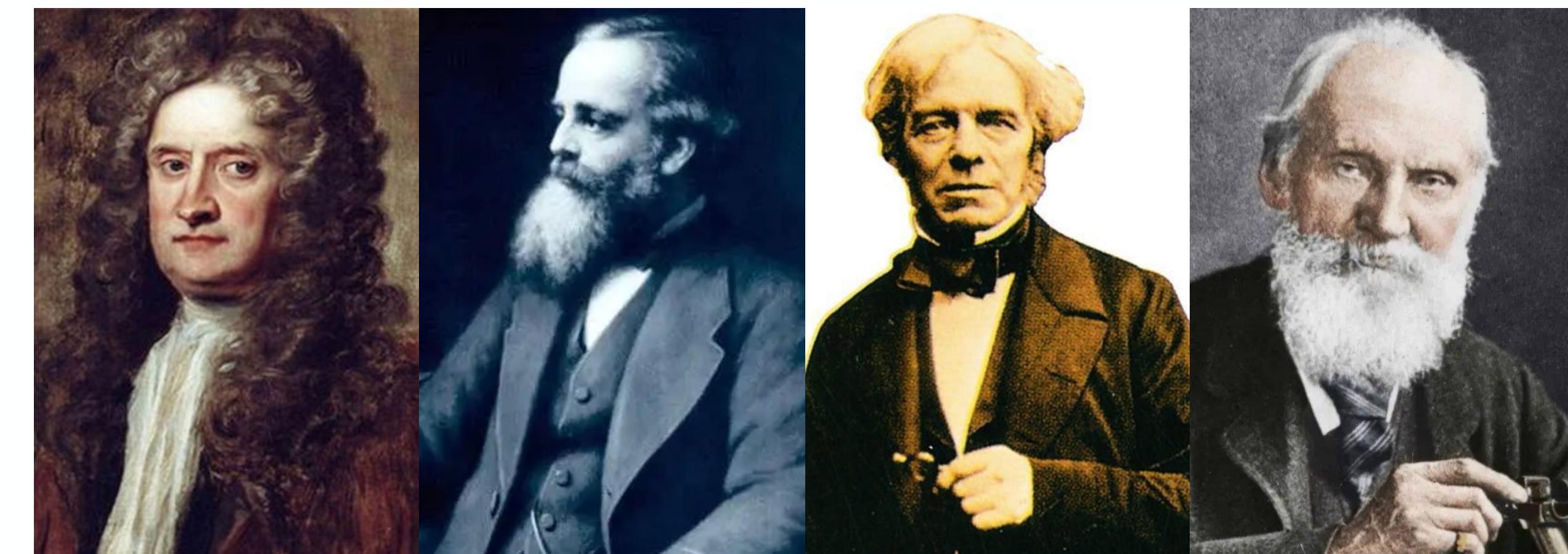
我们知道，蛋白质加热后可以变性，那么，熟鸡蛋通过100℃开水煮20分钟，整体上鸡蛋内容物均有液态变成固态，在返生过程中，不添加任何化学物质，不进行任何物理处理，如加温或者降温，电离辐射等。这是为什么？

鸡卵细胞的由卵黄膜、卵黄和胚盘等组成，卵细胞是否变性，如果变性，即使返生为液态，也难以孵化成雏鸡。这是为什么？欢迎讨论问题如下：

## 文本、字体、排版特征

但是应该很少有人知道，身为科学家的焦耳其实信奉神。

事实上，很多大名鼎鼎的“老”英国科学家都信神，比如近代物理学之父牛顿、统计物理学的奠基人之一麦克斯韦、电学之父和交流电之父法拉第、热力学之父开尔文勋爵……



好吧，超模君还是承认好了，实在不知道怎么写开头，就给大家科普了上面这个小知识。

其他人都拉来活跃气氛的，只有詹姆斯·普雷斯科特·焦耳（James Prescott Joule）才是今天的主人公。

## 酿酒小子

ByteDance 字节跳动

除了信奉神，焦耳身上还有两大标签：物理学家和英国皇家学会会员。

# 伪科学内容识别 文本二分类

## 写作风格

教育教学

写真地理

2020年6月 第22期

### 熟鸡蛋变成生鸡蛋(鸡蛋返生)——孵化雏鸡的实验报告

郭平 白卫云

(郑州市春霖职业培训学校 河南 郑州 450000)

**摘要:**“鸡蛋返生”，顾名思义，就是由熟鸡蛋再变成生鸡蛋。这是一个难以想象的，甚至是不可能的，但是这样奇特的现象确实在郑州春霖职业培训学校发生了。一群特别培训的学生，在郭平老师指导下，正在进行一个奇特实验，即熟鸡蛋重新变成生鸡蛋，并将返生后的生鸡蛋进行孵化或雌雄。并且已经成功返生了40多枚。

**关键词:**生鸡蛋；熟鸡蛋；鸡蛋返生；孵化雏鸡；实验报告

【中图分类号】S831

【文献标识码】A

鸡蛋奇特返生的现象，根据鸡蛋的组织结构及功能。鸡蛋经过高温100℃开水煮20分钟，变成熟鸡蛋后，学生们运用自己的超心理意识能量方法等，将这些熟鸡蛋变成生鸡蛋，现在我们将这种奇特现象分享给科学探秘爱好者，共同探究其内在理论依据。

**实验材料:**鸡蛋10枚，一次性纸质茶杯10个。

**实验场地:**郑州春霖学校507教室。

**实验时间:**2020年6月12日11时。

**室内温度:**摄氏25℃，保持室内安静。

**参加人员:**郑州春霖学校特训生10人，见表1。

表1 观察见证专家及学生家长

姓名	职务(单位)	备注
郭平	校长	郑州市春霖职业培训学校
白卫云	教授	香港毕业大学书记
高建伟	校长	郑州大学设计院系分院院长
单松柏	原国家地质局郑州地质中心主任	地质家单松柏
白立忠	主任医师	河南中医药高等专科学校附属医院
尹杰		山西中医研究所
高爱华	家长	春霖学员 唐晓晓(主要组织者)
陈静霞	家长	孙嘉宝母亲
赵真真	家长	陈敬墨母亲
郭大英	教师	新郑市第二实验小学教师



鸡蛋矿物质组成。对鸡蛋内容物起保护作用。①外层卵壳膜：是层无结构纤维膜，主要是保护鸡蛋内容物水分不丢失。②内层卵壳膜：是一种可透气膜，空气可以进出。③气室：位于鸡蛋的钝端内(在大头与卵壳膜之间)，是由两层卵壳膜之间窄分开形成一个小气室，贮存空气，具有胚胎发育时供应其呼吸功能<sup>[1]</sup>。④系带：卵黄的两端由浓稠的蛋白组成卵黄系带，其功能是维持卵细胞共定于蛋白中心位置，对卵细胞具有起着缓冲作用，可防止卵细胞的震荡。⑤卵黄膜：位于卵白与卵黄之间的一层薄膜，是卵细胞的组成部分。⑥卵黄：呈黄色，位于细胞的中央，是鸡卵胚胎发育的主要营养物质。⑦胚盘：位于卵黄表面中央有一圆形盘状小白点，呈椭球状等。⑧卵白：卵壳膜与卵黄膜之间，为胚胎发育提供水和营养物质。卵黄表面中央有一圆盘状的小白点(就是在蛋黄上看到的小白点)称为胚盘，里面含有细胞核，未受精的卵，胚盘色淡而小，已受精的卵，胚盘色浓而略大，这是因为胚胎发育已已经开始<sup>[2]</sup>。如果此受精卵，胚盘在适宜的条件下就能孵化出雏鸡，胚盘进行胚胎发育的部位<sup>[3-5]</sup>。

我们知道，蛋白质加热后可以变性，那么，熟鸡蛋经过100℃开水煮20分钟，整体上鸡蛋内容物均有液态变成固态，在返生过程中，不添加任何化学物质，不进行任何物理处理，如加温或者降温，电离辐射等。这是为什么？

鸡卵细胞的由卵黄膜、卵黄和胚盘等组成，卵细胞是否变性，如果变性，即使返生为液态，也难以孵化成雏鸡。这是为什么？欢迎讨论问题如下：

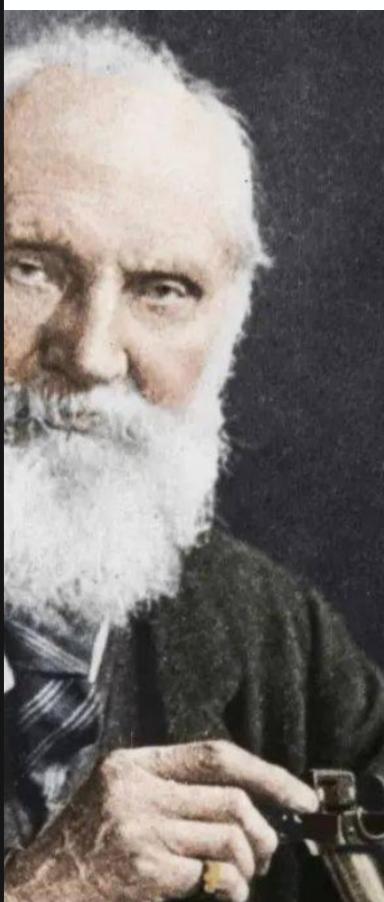
## 文本、字体、排版特征

但是应该很少有人知道，身为科学家的焦耳其实信奉神。

事实上，很多大名鼎鼎的“老”英国科学家都信神，比如近代物理学之父牛顿、统计物理学的奠基人之一麦克斯韦、电学之父和交流电之父法拉第、热力学之父开尔文勋爵



但是应该很少有人知道，**身为科学家的焦耳其实信奉神**。事实上，很多大名鼎鼎的“老”英国科学家都信神，比如近代物理学之父牛顿、统计物理学的奠基人之一麦克斯韦、电学之父和交流电之父法拉第、热力学之父开尔文勋爵……**好吧，超模君还是承认好了，实在不知道怎么写开头，就给大家科普了上面这个小知识。其他人都是拉来活跃气氛的，只有詹姆斯·普雷斯科特·焦耳 (James Prescott Joule) 才是今天的主人公。**除了信奉神，焦耳身上还有两大标签：物理学家和英国皇家学会会员。



好吧，超模君还是承认好了，实在不知道怎么与开头，就给大家科普了上面这个小知识。

其他人都是拉来活跃气氛的，只有詹姆斯·普雷斯科特·焦耳 (James Prescott Joule) 才是今天的主人公。

## 酿酒小子

除了信奉神，焦耳身上还有两大标签：物理学家和英国皇家学会会员。

ByteDance 字节跳动

# 伪科学内容识别 文本二分类

## 存在问题及解决方案

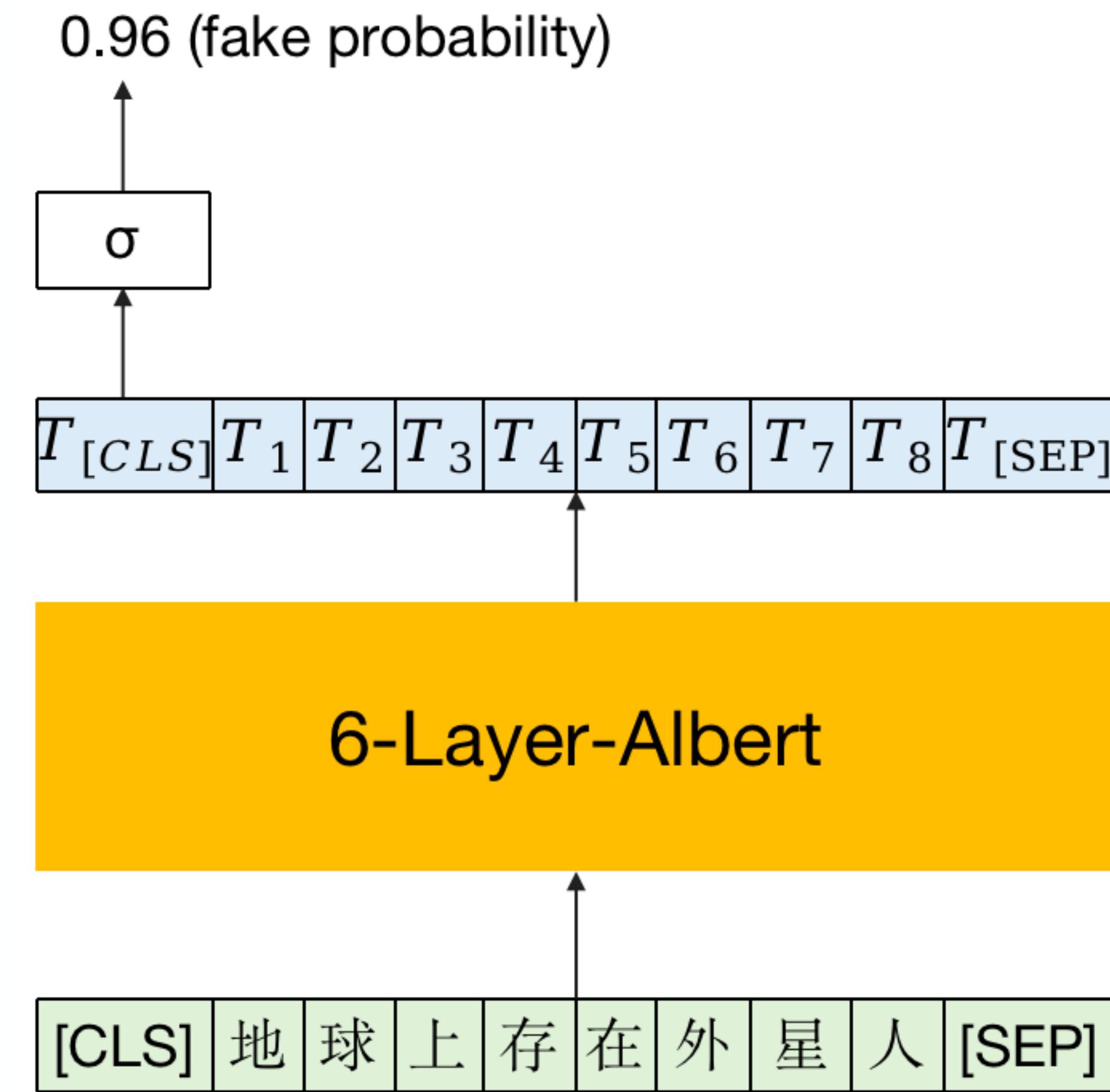
- 伪科学的作者会根据打压情况不断调整自己的写作风格，因此需要不断获取新的数据更新模型
- 之前的训练数据距今时间较长，对于新的写作套路存在召回率和准确率下降的情况，因此需要收集新的数据进行训练。未来希望能形成自动化数据收集及定期训练和改进模型。

## 数据来源

- 头条：图文、微头条、问答
- 抖音：小视频标题以及通过ocr和asr获取的文字信息
- 西瓜：中视频标题以及通过ocr和asr获取的文字信息

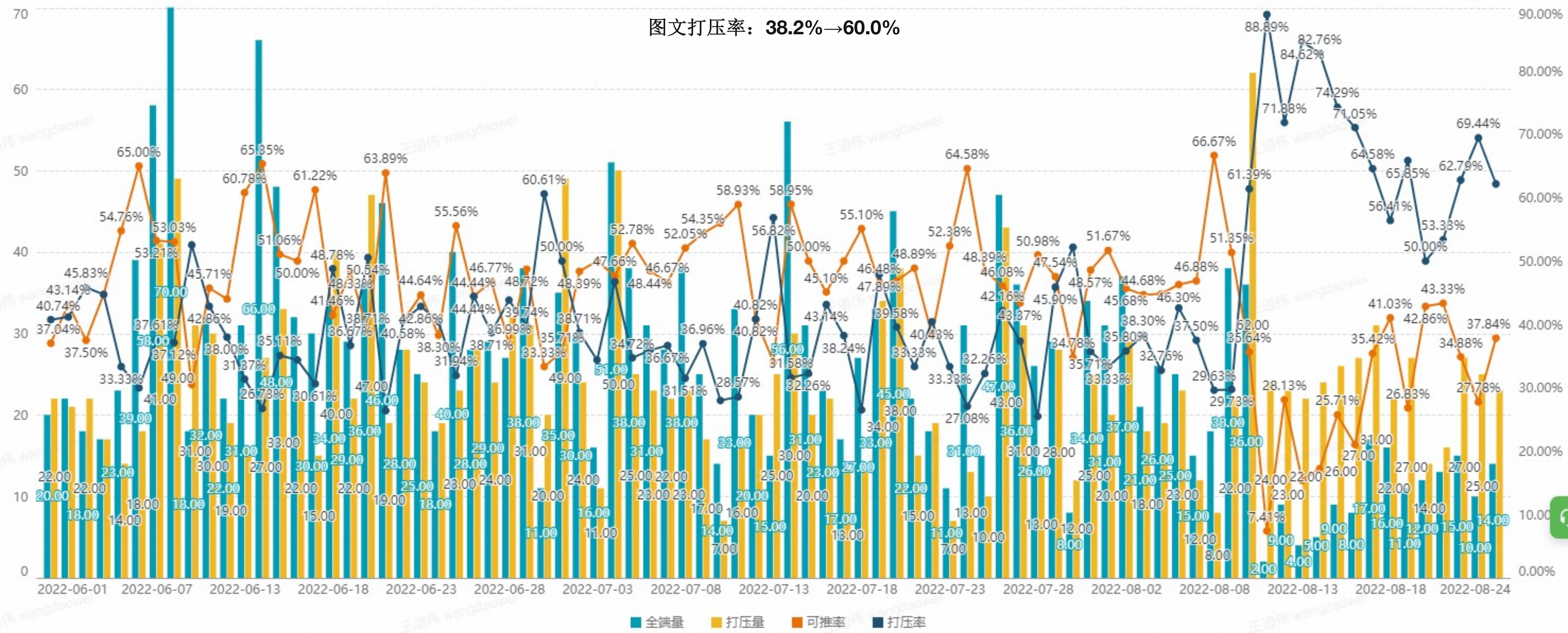
## 现有模型的更新

- 数据抓取和清洗：使用SQL在数据库中抓取数据，筛选出人工审核后进行打压的数据作为正例，不打压的数据作为负例。
- 数据增强：由于正例过少，采用正利double，负例亚采样的方式，保证正负例比例在1:5左右。
- 构造数据集：按时间由远及近分别构造训练集、验证集和测试集，比例大约为10:1:1。
- 模型训练及验证：训练模型，使用验证集选取checkpoint并在测试集上进行测试。
- 人工检验模型效果：将模型命中的数据进行人工标注，验证模型的打压率。



# 伪科学内容识别 文本二分类

图文打压率: 38.2%→60.0%



# 伪科学内容识别 文本二分类



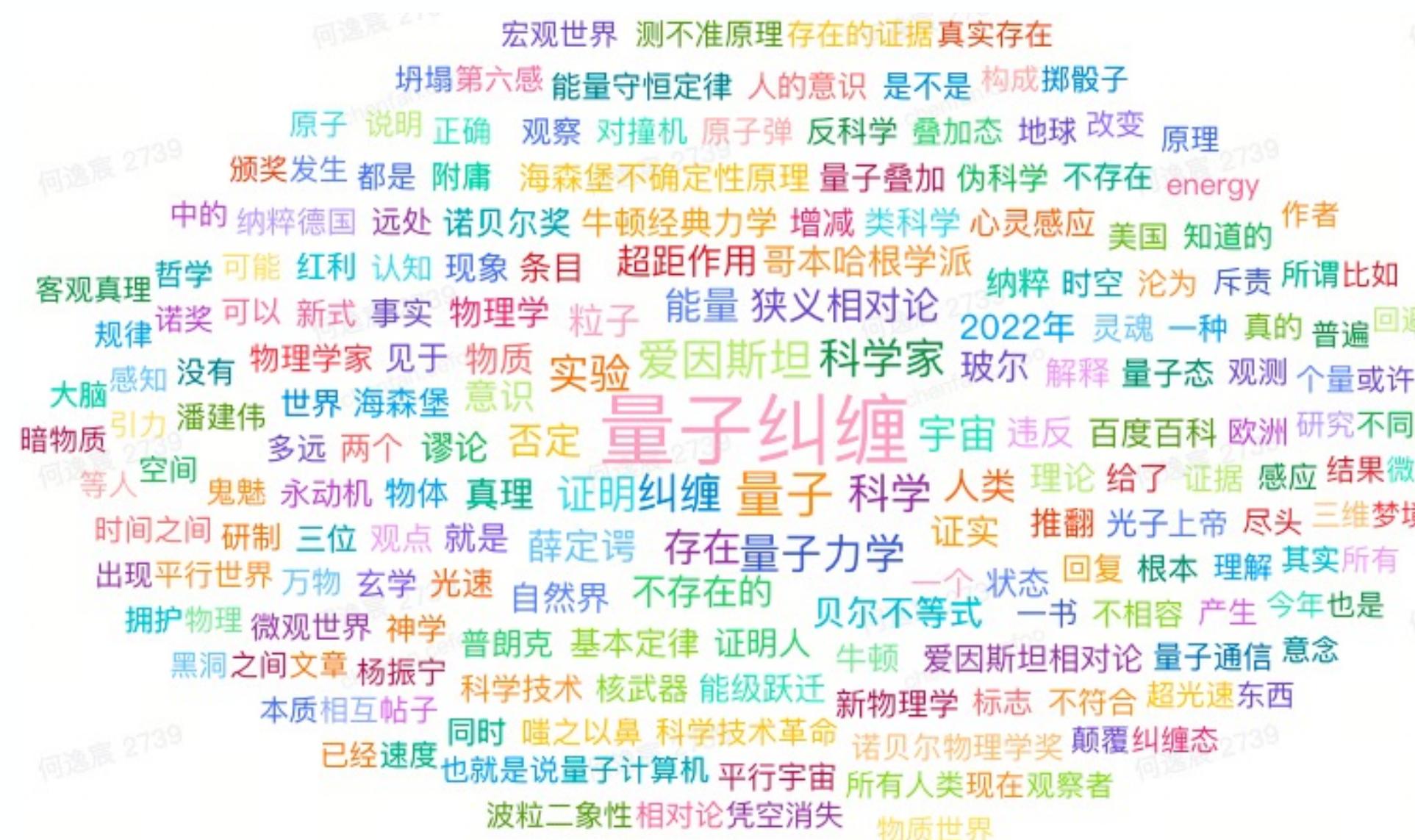
# 伪科学内容识别 文本二分类

## 结果和收益

1. 对于四体裁伪科学模型进行模型训练，打压率由30%提升至40%
  2. 接入了新的中视频体裁，丰富了项目拦截范围，目前中视频周审出1000+，打压率超40%
  3. 伪科学周均拦截科学虚假内容700+，年度拦截伪科学虚假内容3.5W+
  4. 对全局内容的调节，全局科学虚假内容占比图文三体裁从2022年3月的6.2%虚假占比，降低到了3.8%

## 实现模型自动更新

将上述更新模型流程自动化，自动抓取数据，进行数据增强，训练模型，在相同的测试集下测试新旧模型的效果，效果有提升则自动更新线上模型。



其实这个话题，目前来说也没有哪个科学家能给出一个确切的结论说金字塔到底是不是外星人建的，但是也不能排除这种可能。即使到现在金字塔还是有很多的未解之谜，即使用现在的技术去

# 信源库系统搭建

# 项目背景

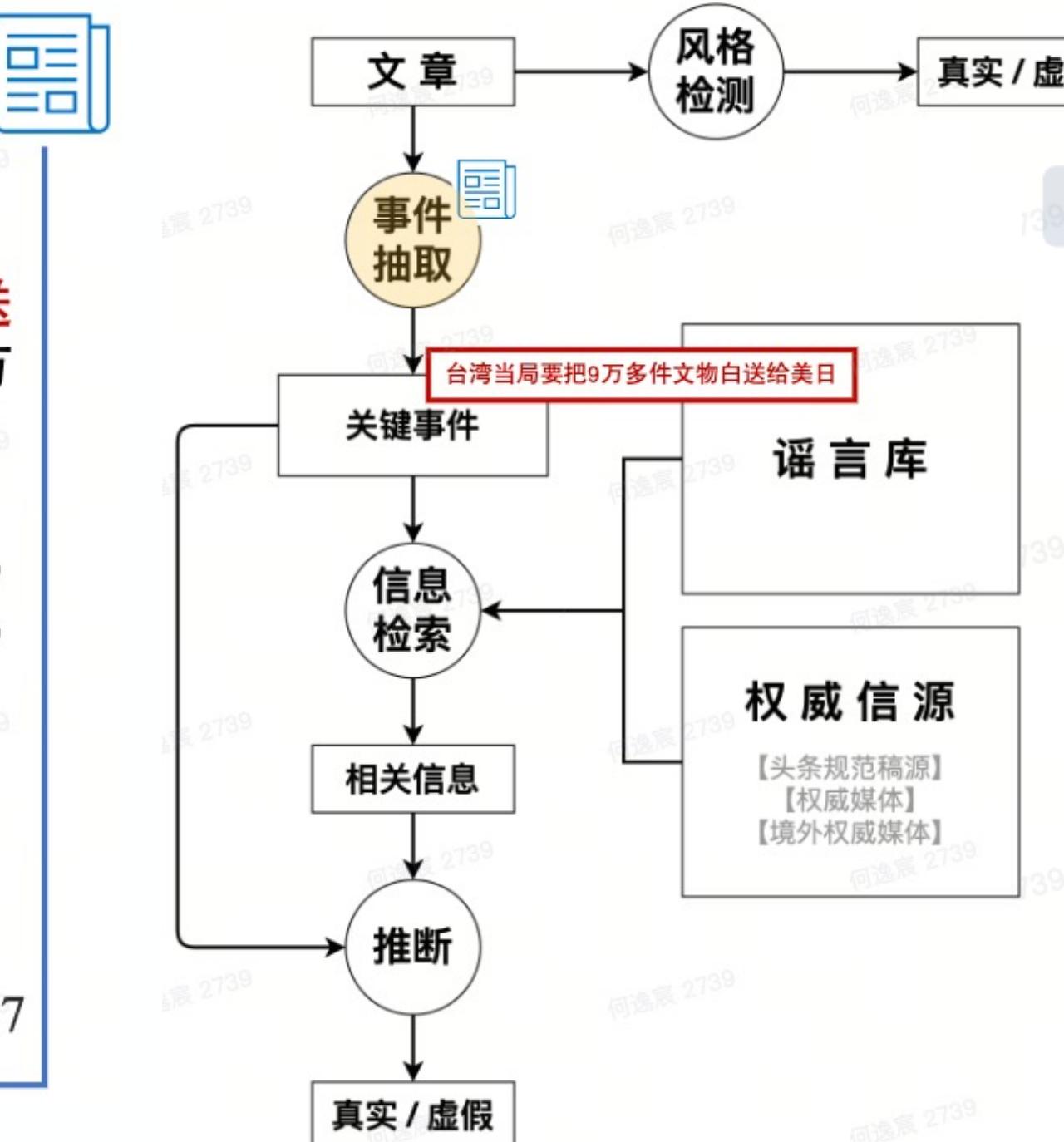
## 白名单-信源库

权威信源库：覆盖国内外权威信源发布信息。

- 头条站内规范稿源
- 权威媒体（站点150+）
  - 新华社 / 光明网 / 央视网 / ...
  - 政府机构 / 公司官网 / ...
  - 垂类权威媒体
  - 路透社 / CNN / 法新社 / 半岛电视台 / 俄罗斯卫星通讯社 / ...
  - 收录文章数400w+, 日均新增5k+

最近台湾局势非常的紧张，台湾方面已经乱了，最近有一个来自海外网的消息称：**台湾当局要把9万多件文物白送给美日！**要知道每件文物的价值都在9万美金以上，有些甚至要数千万数亿万的国宝。  
表面上这是一次紧急的疏散的演练，但是实际是在为这些文物想到转移方案，自从佩洛西来台湾后，台湾的局势发生了很大的变化，一旦发生不测，这些文物就有可能落到大陆的手中。  
但是关于这批文物的细节和内容，相关人员却一直没有说。

2022-08-08 17:16:57



## 黑名单-谣言库

包含谣言3w+, 周新增200+

- 碱性食物能预防头皮屑
- 凯里剑河县交警收黑钱
- 雷锋被从教材中删减
- 山楂能降血脂

虚 假 台湾当局要把9万多件文物白送给美日, 2022-08-08 17:16:57

- 反对** **台湾当局要把9万多件文物白送给美日**
- 南方plus客户端: 台北故宫辟谣台当局准备将文物转移至美日：“绝无此事”， 2022-08-08 17:14:00
  - 澎湃新闻: 岛内传台当局要将故宫文物送美日“保护”，台北故宫方面支吾， 2022-08-08 16:11:24
  - 海峡网: 传民进党当局准备将9万藏品转移美日寻求“保护”，台北故宫回应了， 2022-08-08 15:04:39
  - .....

虚假模型信息	
文中短句1:	<a href="#">台湾当局要把9万多件文物白送给美日</a>
判断结果:	与权威信源矛盾，疑似虚假
命中权威信源信息:	<a href="#">台北故宫辟谣台当局准备将文物转移至美日：“绝无此事”</a>
南方plus客户端	2022-08-08 17:14:00
岛内传台当局要将故宫文物送美日“保护”、台北故宫方面支吾	<a href="#">澎湃新闻</a> 2022-08-08 16:11:24
9万藏品将转移美日？台北故宫回应	<a href="#">鲁网</a> 2022-08-08 16:08:30
传民进党当局准备将9万藏品转移美日寻求“保护”、台北故宫回应了	<a href="#">海峡网</a> 2022-08-08 15:04:39
台北故宫博物院回应：绝无此事	<a href="#">西安晚报</a> 2022-08-08 13:43:28

# 项目背景

## 白名单-信源库

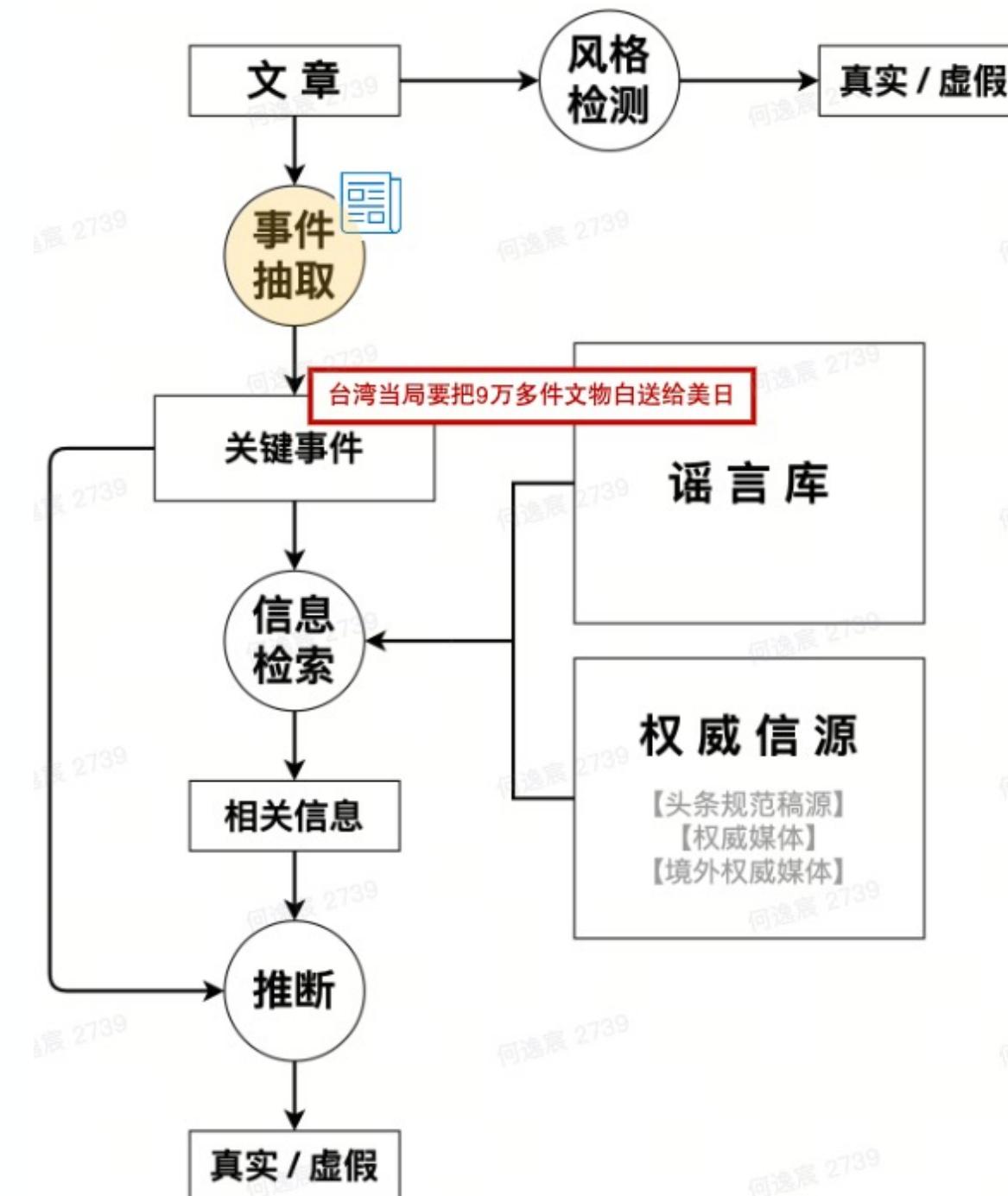
权威信源库：覆盖国内外权威信源发布信息。

- 头条站内规范稿源
- 权威媒体（站点150+）
  - 新华社 / 光明网 / 央视网 / ...
  - 政府机构 / 公司官网 / ...
  - 垂类权威媒体
  - 路透社 / CNN / 法新社 / 半岛电视台 / 俄罗斯卫星通讯社 / ...
  - 收录文章数400w+, 日均新增5k+

## 黑名单-谣言库

包含谣言3w+, 周新增200+

- 碱性食物能预防头皮屑
- 凯里剑河县交警收黑钱
- 雷锋被从教材中删减
- 山楂能降血脂

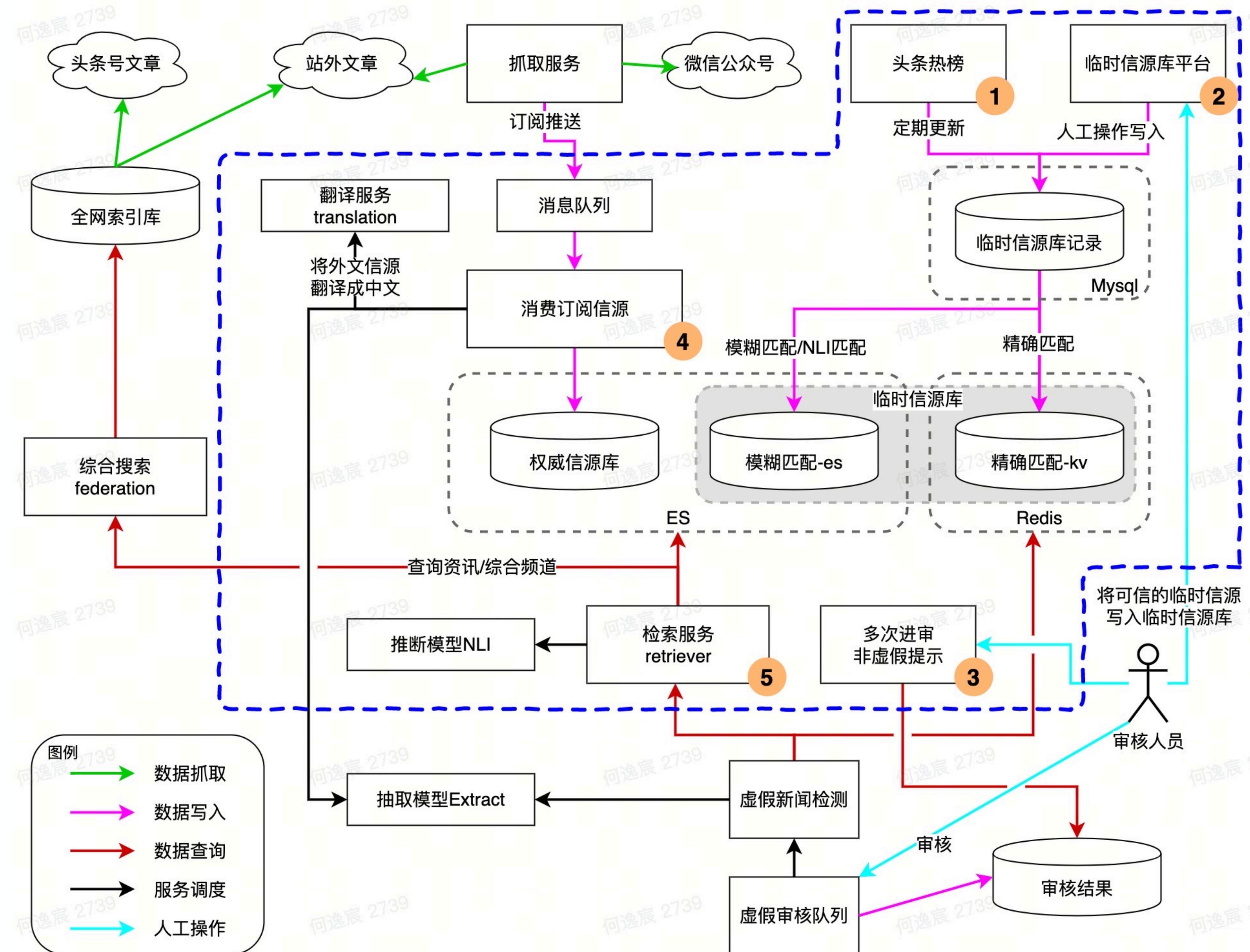


虚假 **适度吃辣椒确实可以减肥**

蕴含 **吃辣椒可以减肥**

• **抗糖水果椒能减肥**

# 信源库系统搭建



# 热榜内容自动入库 Python | Redis | MySQL

微头条有许多发文是针对热榜事件的评论，热榜内容很多是突发的高热度事件，其内容是经过审核的但是没有录入权威信源库，因此会造成大量的没有权威信源的情况。



运筹帷幄荷叶nn

2022-08-21 11:15

+ 关注

虚假模型信息

#赖岳谦评台军与解放军军舰对峙画面# 赖教授说的对！为他点赞！

① 举报

文中短句1：赖岳谦评台军与解放军军舰对峙画面

判断结果：无相关权威信源，疑似虚假

命中权威源 无  
信息：

使用定时job，每隔5min抓取一次热榜内容，并将其加入临时信源库`abase`中，每次加入的内容设置其`ttl`为30min。模型抽取出的内容会与临时信源库进行比对，若命中则不进审，日均减少进审量约100条。

```
client = Client(psm="", table="")
url = "https://api.toutiaoapi.com/hot-event/hot-board/?origin=hot_board"
create_time = int(datetime.now().timestamp()) # 获取当前时间戳
response = requests.get(url) # 发送 HTTP GET 请求
data = json.loads(response.text)["data"] # 解析 JSON 响应
db = connect(db_psm="")
cursor = db.cursor()
for cluster in data:
    title = cluster["Title"].strip() # 提取热榜标题
    client.set(title, "true", ex=1800) # 将标题写入 Bytedatabase
    cursor.execute("INSERT INTO temp_trusted_data(statement, create_time, match_method, ttl, owner, status VALUES (%s, %s, %s, %s, %s, %s) ON DUPLICATE KEY UPDATE ttl = %s", (title, create_time, 1, create_time + 1800, "hot_board", create_time + 1800, 1))
db.commit() # 提交 MySQL 事务
cursor.close() # 关闭数据库连接和游标
db.close()
```

# 临时信源库 Golang | Redis | MySQL | ES

存在许多类似热榜的临时信源，因为热度不够高没能进入热榜，我们将一些多次进审不打压的case视为这种临时信源通过飞书bot反馈给运营同学，在他们人工判断的情况下决定是否将其加入临时信源库(MySQL, 用于前端展示)，并设置精确匹配(Redis)和模糊匹配(ES)。

```
func customizedRegister(r *server.Hertz) {
    r.POST("/add", handler.Add)           // 增
    r.DELETE("/delete", handler.Delete)   // 删
    r.PUT("/update", handler.Update)      // 改
    r.GET("/query", handler.Query)        // 查
}
```

创建

\* 事件内容

匹配方式

创建人

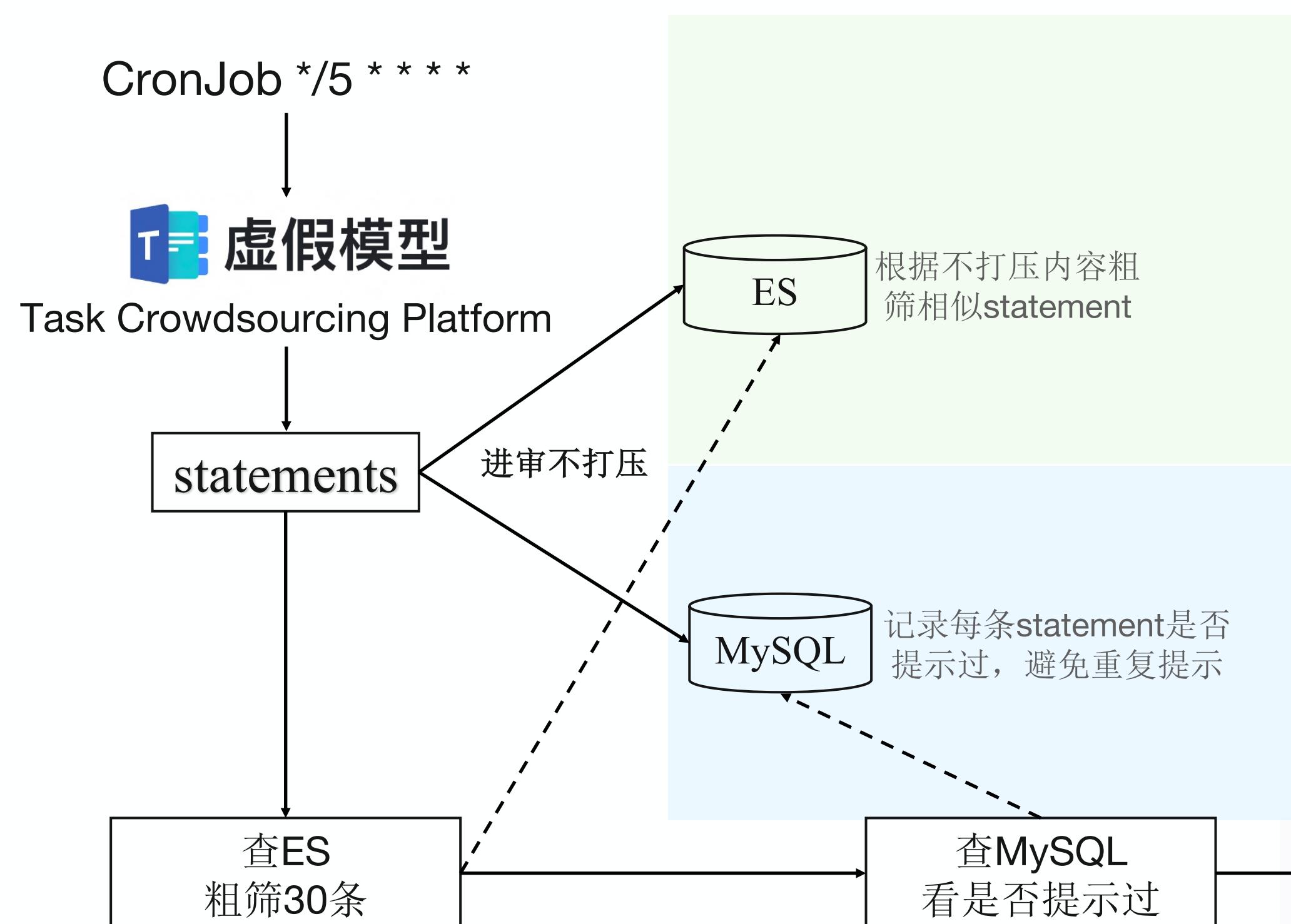
过期时间 (默认一天有效)

输入关键词查找  刷新  luojingshu

ID	事件	全部	匹配模式	状态	过期时间	操作人	最后更新时间	操作
38481	中国电科CETC加班事件	在线	NLU匹配	在线	2023/4/7 00:00:00	luojingshu	2023/4/6 09:30:30	<input type="button" value="编辑"/> <input type="button" value="下架"/>
38479	美国政府周三表示，他们已对格鲁吉亚最高司法委员会的4名成员实施签证制裁，理由是他们涉嫌“严重腐败”		NLU匹配	在线	2023/4/7 00:00:00	luojingshu	2023/4/6 09:22:47	<input type="button" value="编辑"/> <input type="button" value="下架"/>
38457	河北8次地震了		精确匹配	在线	2023/4/7 00:00:00	luojingshu	2023/4/6 07:38:30	<input type="button" value="编辑"/> <input type="button" value="下架"/>
38314	五角大楼的一份报告，中国首次将至少一艘核武装弹道导弹潜艇保持在海上		NLU匹配	在线	2023/4/7 00:00:00	luojingshu	2023/4/5 14:06:23	<input type="button" value="编辑"/> <input type="button" value="下架"/>
38310	西南大学一教授被曝潜规则女博士三年		NLU匹配	过期	2023/4/6 00:00:00	luojingshu	2023/4/5 13:53:03	<input type="button" value="编辑"/> <input type="button" value="下架"/>

共 221 条 < 1 2 3 ... 23 >

# 多次进审非虚假提示 Python | MySQL | ES | CronJob



```

    "gid": {
        "type": "keyword"
    },
    "statement": {
        "fields": {
            "keyword": {
                "ignore_above": 256,
                "type": "keyword"
            }
        },
        "type": "text"
    }
}

```

```

CREATE TABLE `noticed_data` (
    `id` bigint unsigned NOT NULL AUTO_INCREMENT COMMENT '自增ID',
    `statement` varchar(255) NOT NULL COMMENT '事件描述',
    `notice_time` int DEFAULT NULL COMMENT '提醒时间',
    `status` int DEFAULT '1' COMMENT '是否打压',
    `gid` varchar(64) NOT NULL COMMENT '文章id',
    PRIMARY KEY (`id`),
    UNIQUE KEY `uniq_statement_gid` (`statement`,`gid`)
)

```

临时信源 虚假新闻临时信源提示群 16 | 应 2  
 消息 Pin 群公告 云文档 +  
 虚假新闻bot 机器人 | 用于查询虚假新闻以及定位

打压数量: 1  
 美参议员: 如果中国封锁台湾, 美国就封锁中国 [https://gip.bytedance.net/bluewhale/content/article\\_detail?gid=7220444268320686629](https://gip.bytedance.net/bluewhale/content/article_detail?gid=7220444268320686629)

不打压数量: 4  
 美国南卡参议员林赛·格雷厄姆叫嚣: “美国需要中国明白, 如果他们封锁台湾, 美国就会封锁中国 [https://gip.bytedance.net/bluewhale/content/article\\_detail?gid=1762831653118988](https://gip.bytedance.net/bluewhale/content/article_detail?gid=1762831653118988)

美国南卡参议员林赛·格雷厄姆叫嚣: “美国需要中国明白, 如果他们封锁台湾, 美国就会封锁中国 [https://gip.bytedance.net/bluewhale/content/article\\_detail?gid=1762818800403468](https://gip.bytedance.net/bluewhale/content/article_detail?gid=1762818800403468)

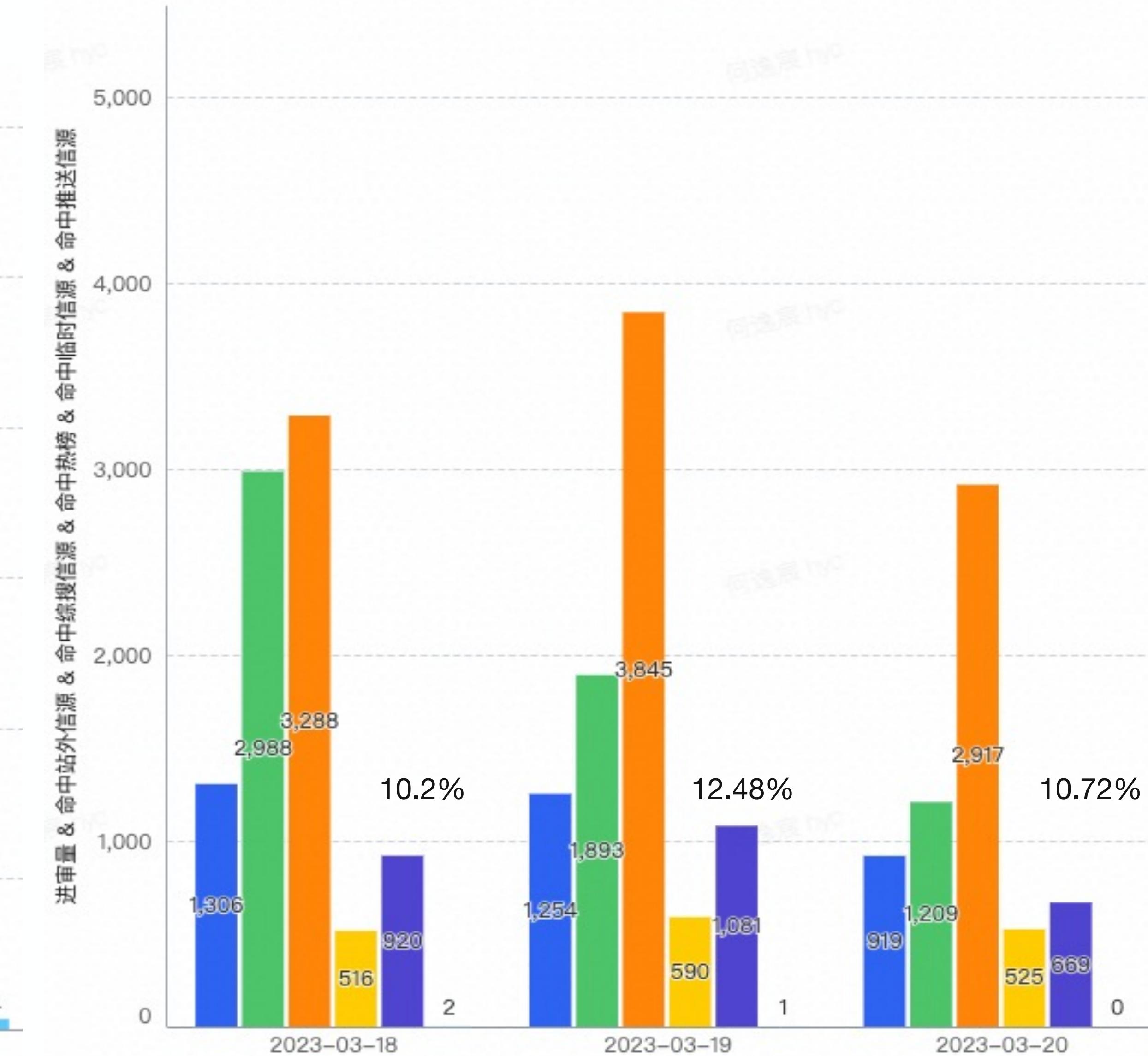
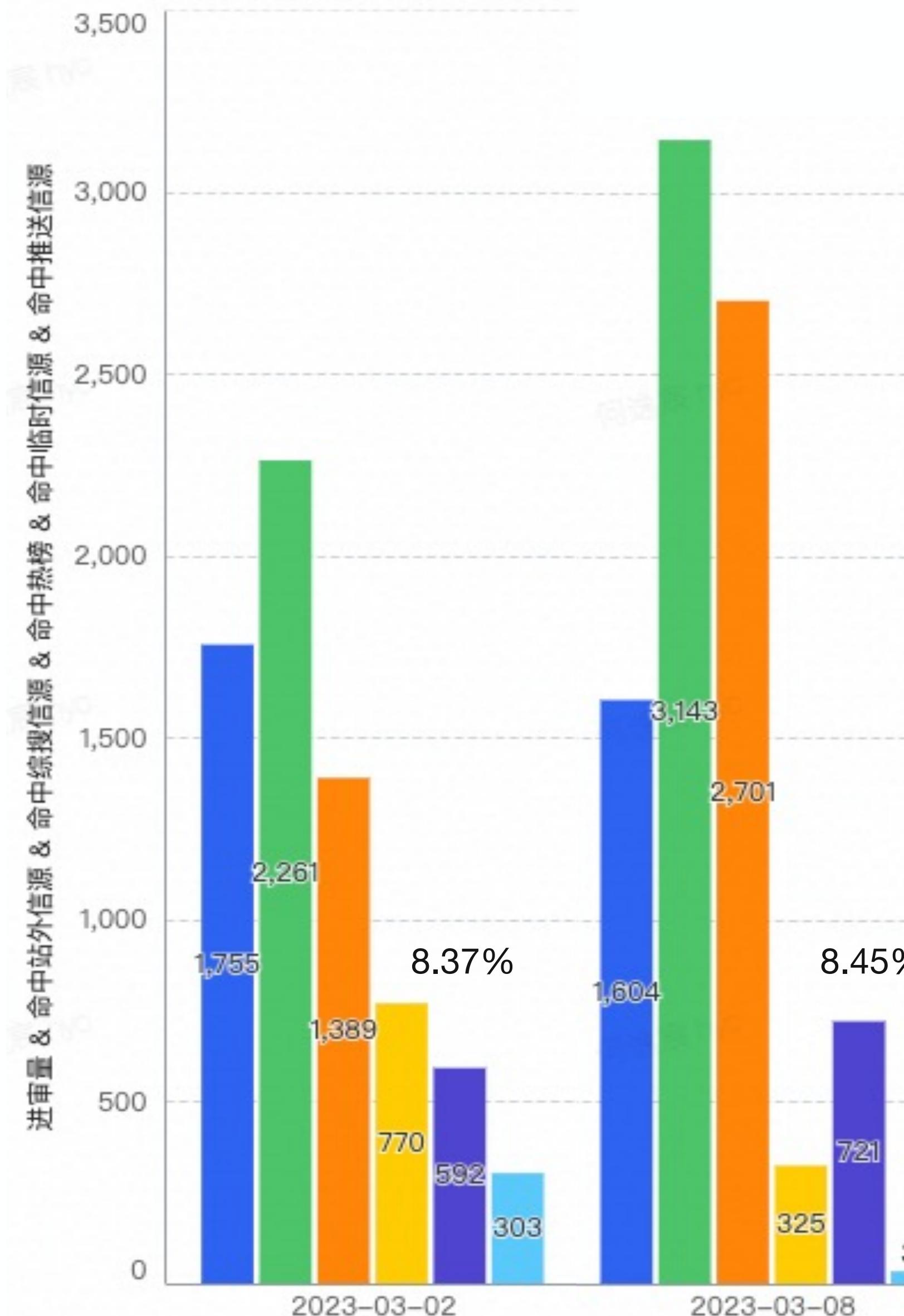
美国南卡参议员林赛·格雷厄姆叫嚣: “美国需要中国明白, 如果他们封锁台湾, 美国就会封锁中国 [https://gip.bytedance.net/bluewhale/content/article\\_detail?gid=1762826507550735](https://gip.bytedance.net/bluewhale/content/article_detail?gid=1762826507550735)

美顶级智库: 中国大陆若武统台湾, 美国就封锁中国海空贸易通道 [https://gip.bytedance.net/bluewhale/content/article\\_detail?gid=716911406](https://gip.bytedance.net/bluewhale/content/article_detail?gid=716911406)

OK | 罗婧姝



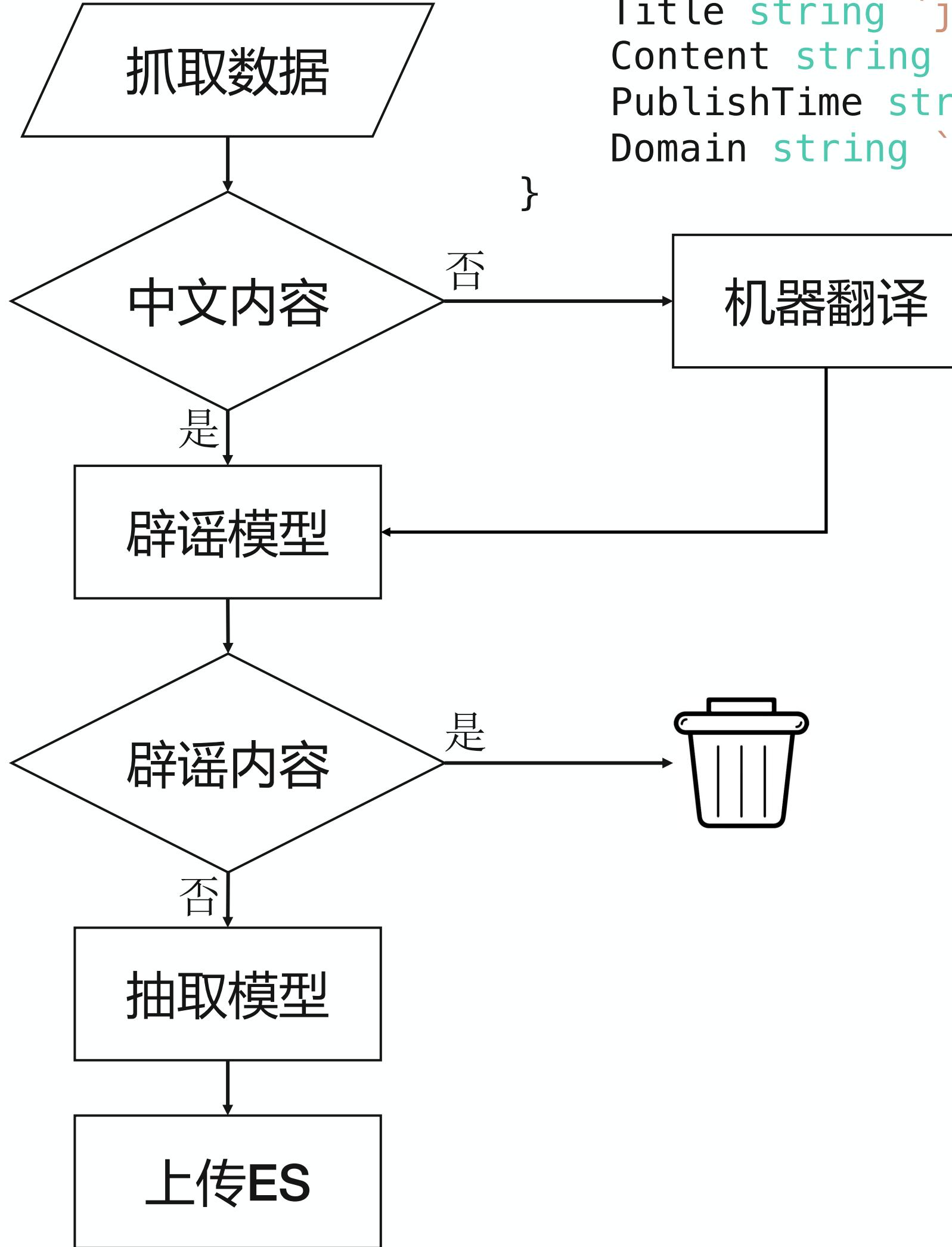
# 多次进审非虚假提示 Python | MySQL | ES | CronJob



进审量  
命中站外信源  
命中综搜信源  
命中热榜  
命中临时信源  
命中推送信源

动

# 订阅信源消费 Golang | ES



```

type Doc struct {
    URL string `json:"url"`
    Title string `json:"title"`
    Content string `json:"content"`
    PublishTime string `json:"publish_time"`
    Domain string `json:"domain"`
}
    
```

```

type ESDoc struct {
    ID uint64 `json:"id"`
    Title string `json:"title"`
    Texts []string `json:"texts"`
    Statements []string `json:"statements"`
    URL string `json:"url"`
    Domain string `json:"domain"`
    PublishTime string `json:"publish_time"`
    UpdateTime string `json:"update_time"`
    OriginalTitle string `json:"original_title"`
    OriginalContent string `json:"original_content"`
    OriginalTexts []string `json:"original_texts"`
}
    
```

Annotations for the ESDoc struct fields:

- ID uint64 `json:"id"` // Translated Title
- Title string `json:"title"` // Translated Texts
- Texts []string `json:"texts"` // 抽取模型抽到的关键句 \*
- Statements []string `json:"statements"` // Doc.URL
- URL string `json:"url"` // Doc.Domain
- Domain string `json:"domain"` // Doc.PublishTime
- PublishTime string `json:"publish\_time"` // time.Now()
- UpdateTime string `json:"update\_time"` // Doc.Title
- OriginalTitle string `json:"original\_title"` // Doc.Content
- OriginalContent string `json:"original\_content"` // Content去掉html标签并切句
- OriginalTexts []string `json:"original\_texts"` // Content去掉html标签并切句

Partition ID	Latest Offset	Earliest Offset	lag	Group offset
0	2305803	2296315	0	2305799
1	2304876	2295011	0	2304876
2	2307851	2287276	0	2307849
3	2304542	2286630	0	2304541
4	2306233	2287978	0	2306229

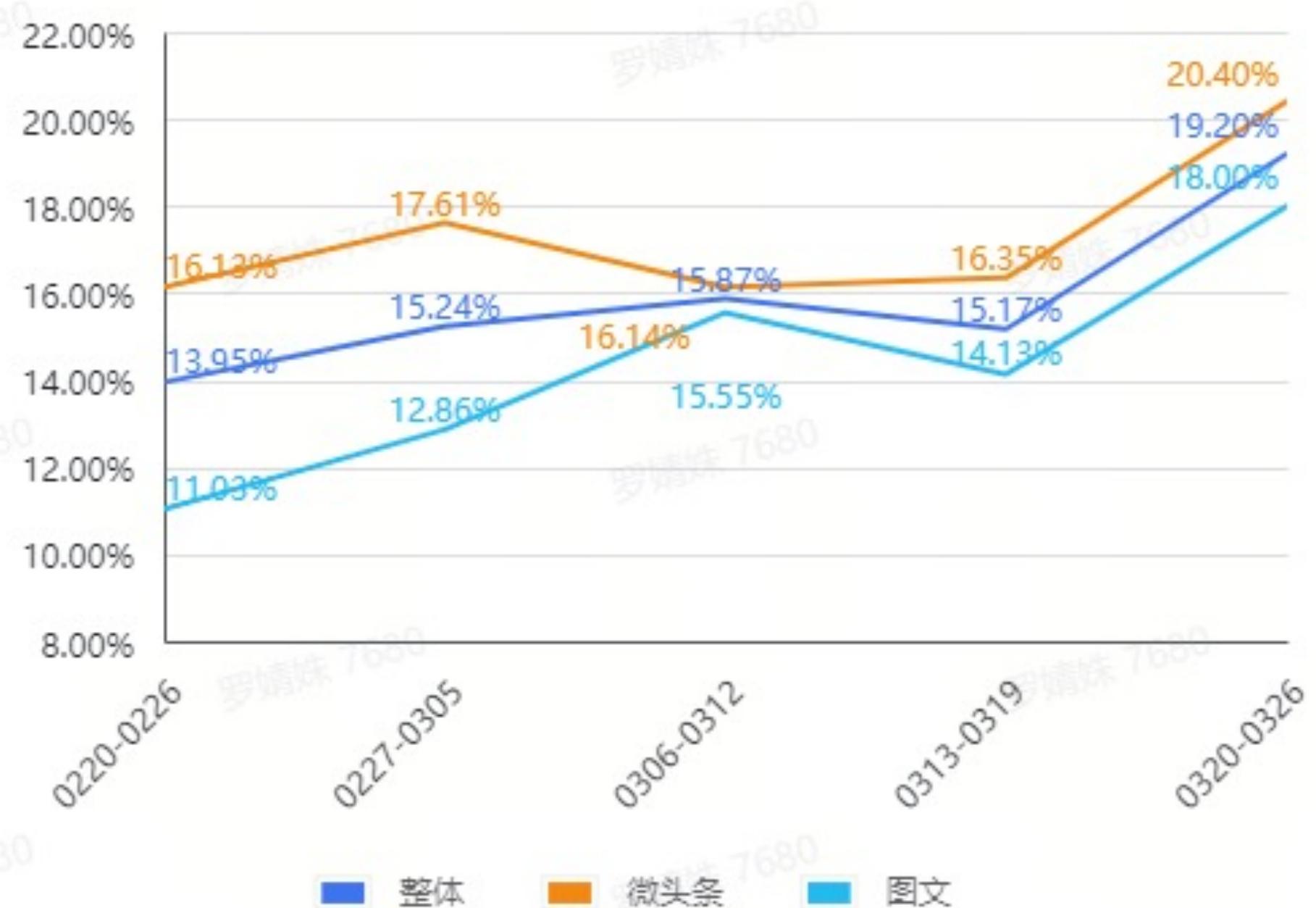
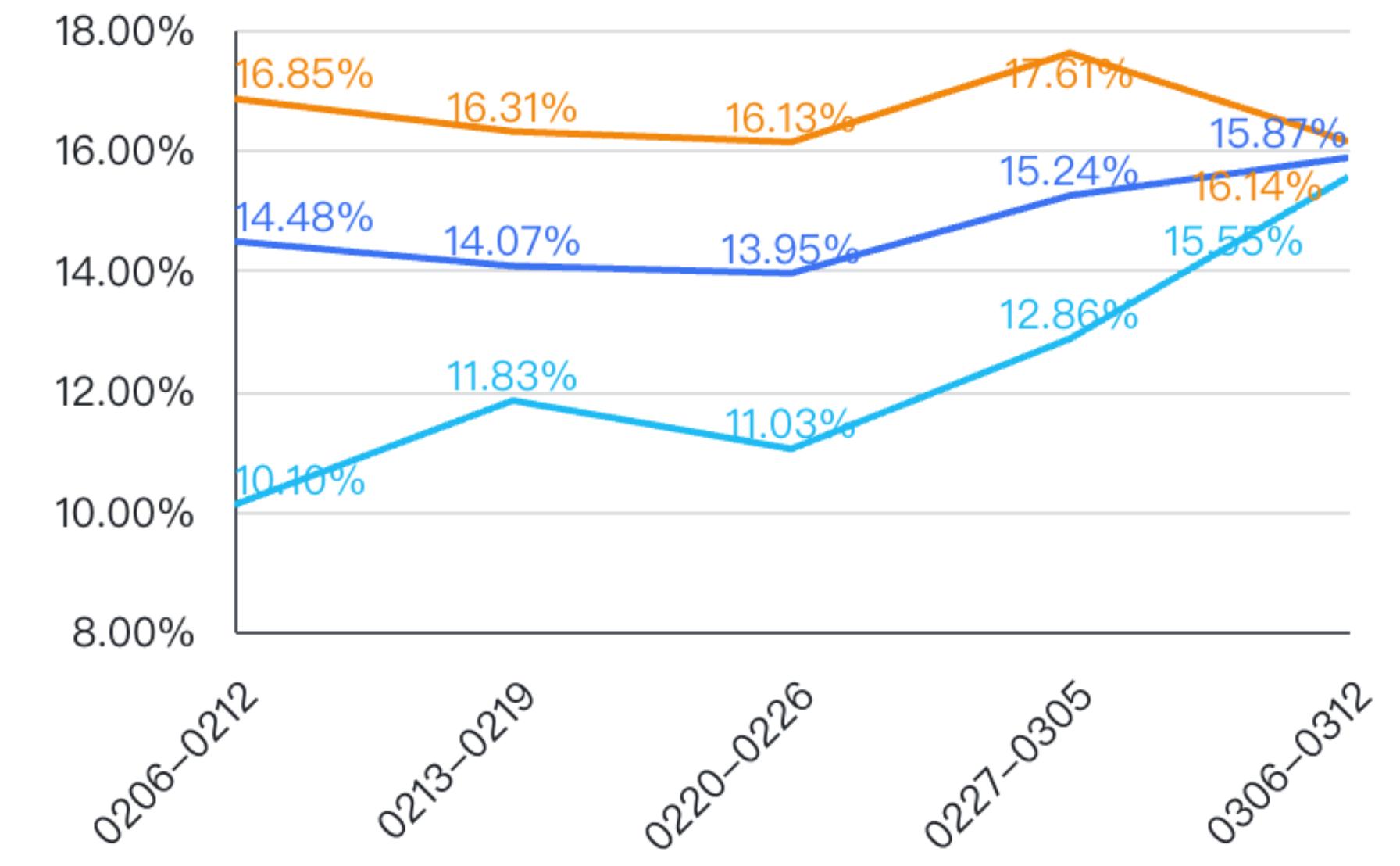
# 检索服务 Golang

改进服务，为检索到的信源计算nli分数并返回

query: 12月3日，国家宗教事务局令第17号公布了《互联网宗教信息服务管理办法}

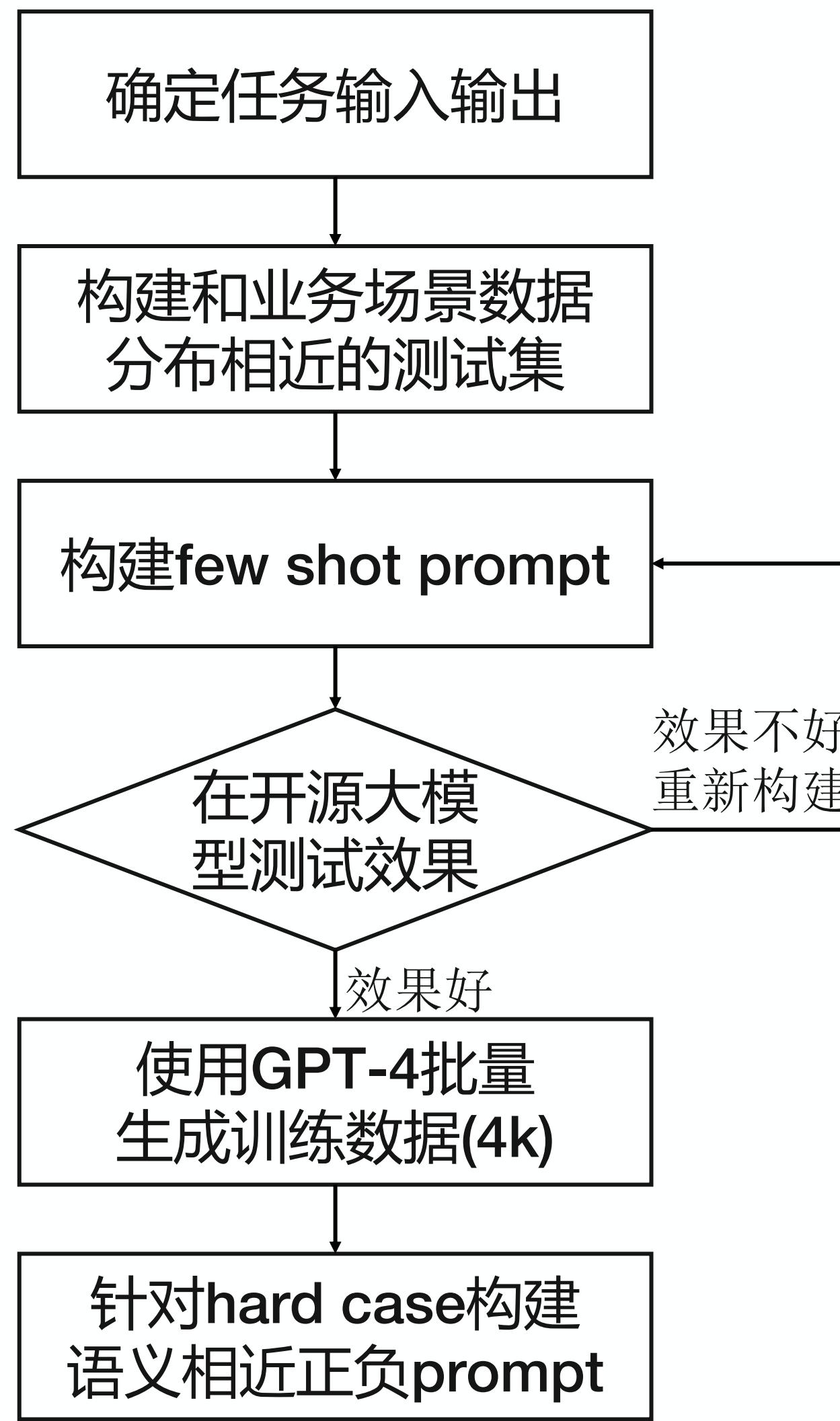
```
{  
    "nli_scores": [  
        {  
            "pred": 2,  
            "prob": [  
                0.14509249,  
                0.010170289,  
                0.46734226,  
                0.000067420835,  
                0.3773276  
            ],  
            "statement": "《互联网宗教信息服务管理办法》全文来了"  
        }, ...  
    ],  
    "publish_time": 1680850399,  
    "rel_score": 0.5389053225517273,  
    "sitename": "",  
    "statements": [  
        "《互联网宗教信息服务管理办法》全文来了", ...  
    ],  
    "title": "《互联网宗教信息服务管理办法》全文来了！学起来！",  
}
```

虚假打压率



# 大模型训练

# 谣言库 GPT4 - LLaMA 业务场景大模型训练流程



【文章】：生姜还有杀灭口腔致病菌和肠道致病菌的作用，用生姜水含漱治疗口臭和牙周炎，有一定的疗效。大院姜农余丰收正在收获成熟的生姜 六、白姜的吃法和注意事项 铜陵白姜吃法很多，有喝姜汤（姜汁），吃姜粥等，糖冰姜、姜茶、姜炒制菜肴的食用更是司空见惯。既能使味道鲜美，又有助于开胃健脾，促进食欲，帮助消化。

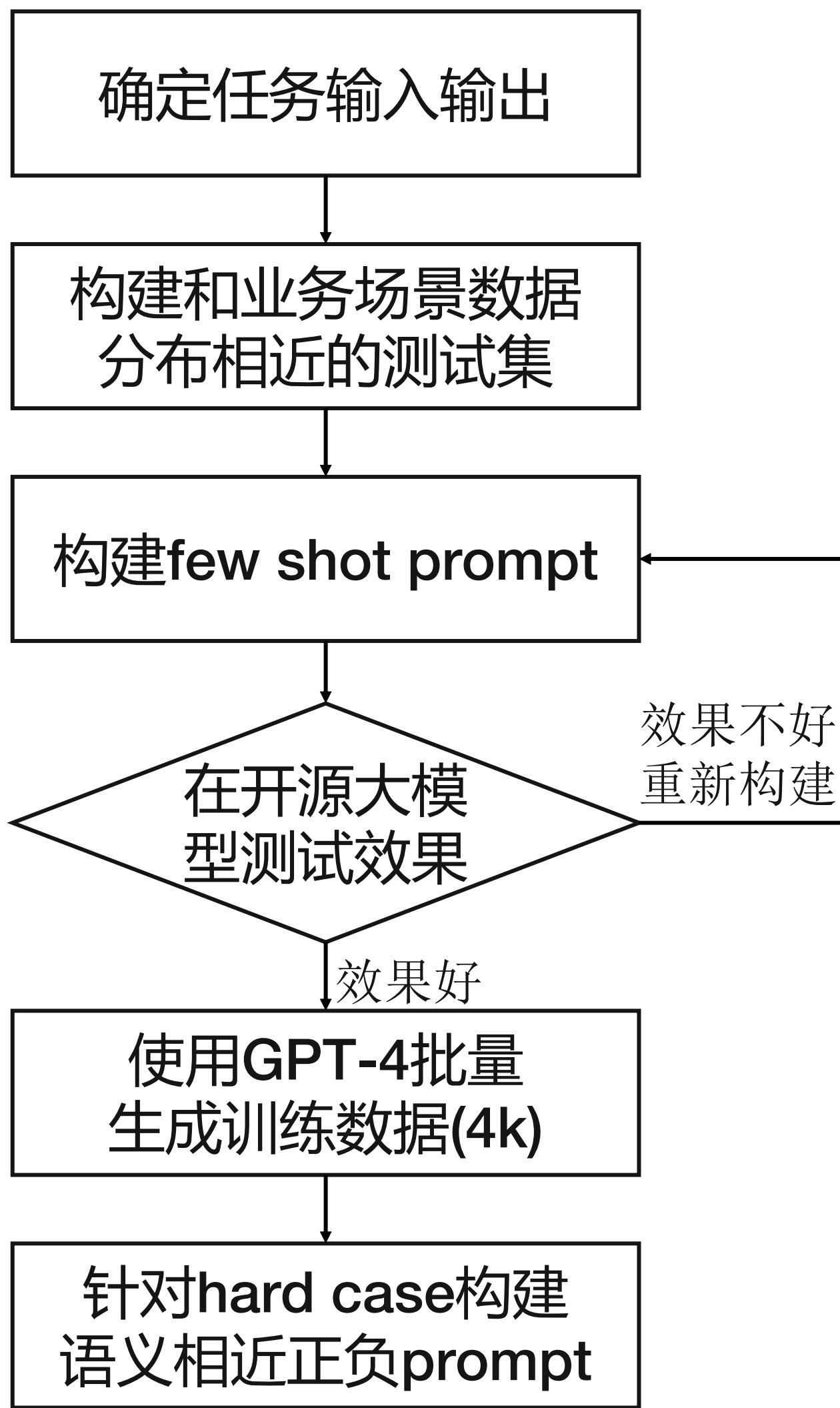
【谣言点】：生姜水漱口治疗牙周炎

判断依据：

1. 【文章】提到用生姜水含漱治疗口臭和牙周炎，有一定的疗效。
2. 【谣言点】提到生姜水漱口治疗牙周炎。
3. 用生姜水含漱治疗口臭和牙周炎可以视为生姜水漱口治疗牙周炎的具体应用。
4. 【文章】可以推出【谣言点】，是。

请判断【文章】是否蕴含【谣言点】，请回答是或否，并说明理由

# 谣言库 GPT4 - LLaMA



【文章】事业单位迎来动态普调，有42种事业编制被取消，改为“聘任制”，这些事业单位的工作人员将与单位签订劳动合同，合同到期之后，根据双方意愿选择是否继续续签。其实这样的改革挺好，取消事业编制，打破铁饭碗，才能服务好企业和群众，促进经济快速发展。取消事业编制，也能把那些靠关系进编制的人员裁掉，还一个公平竞争的环境，有能力的继续升级，混日子的就只能卷铺盖走人了。

【谣言点】国家全面取消事业编制。

判断依据：

1. 【文章】提到有42种事业编制被取消。
2. 【谣言点】提到国家将全面取消事业编制。
3. 42种事业编制被取消不能说明全面取消事业编制。
4. 【文章】不能推出【谣言点】，否。

【文章】一个人有几斤几两，大概率上年岁越大越心里有数，现在科学对算命有一种解释那就是宇宙磁场原理，听起来就容易理解了，你和你周围的磁场互为环境，如果有能力去吸引或排斥掉相应的磁场，那一定程度上可以做到趋吉避凶。但这里有句俗话说“一个人的命运七分天注定，三分靠打拼”，我以为打拼并不是肤浅的理解的好好学习、好好工作，打拼不单是对自己的负责，也是一种不断利他的过程，一个心存善念，不断回报社会的人，他的命运一般都不会差

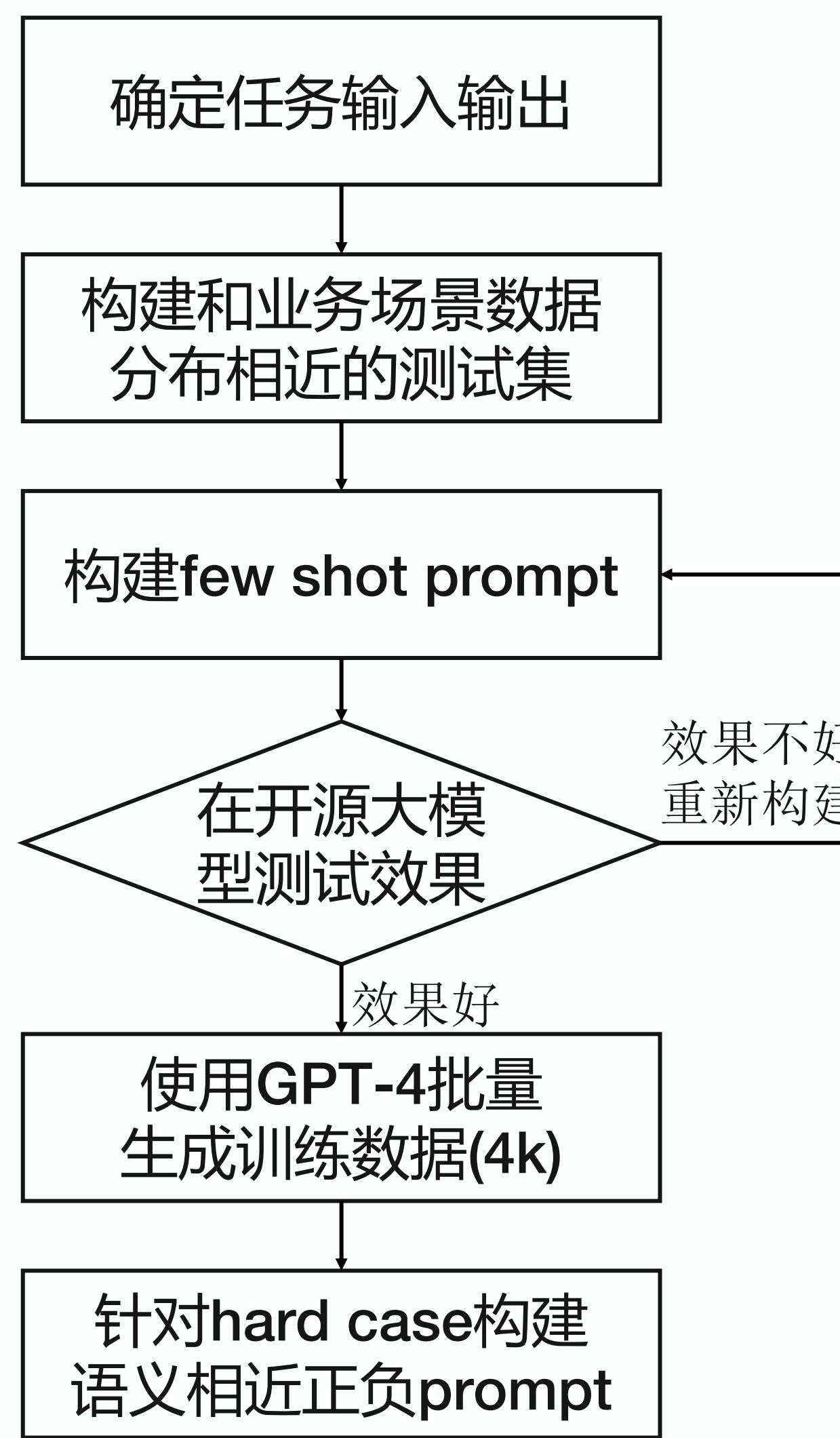
【谣言点】用科学解释算命

判断依据：

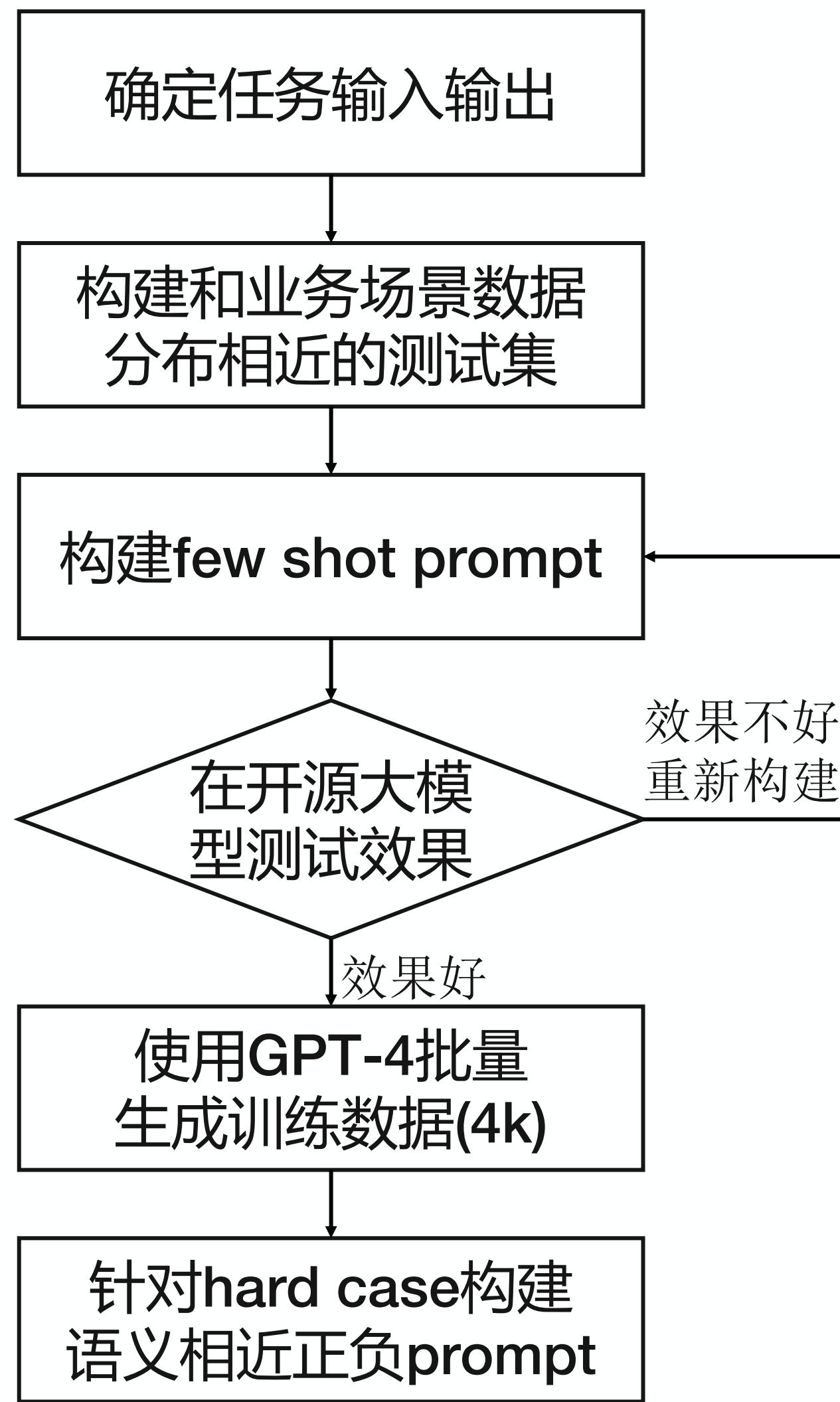
1. 【文章】提到宇宙磁场原理是科学对算命的一种解释。
2. 【谣言点】提到可以用科学解释算命。
3. 宇宙磁场原理是用科学解释算命。
4. 【文章】可以推出【谣言点】，是。

【文章】{}  
【谣言点】{}  
判断依据：

# 谣言库 GPT4 - LLaMA 业务场景大模型训练流程



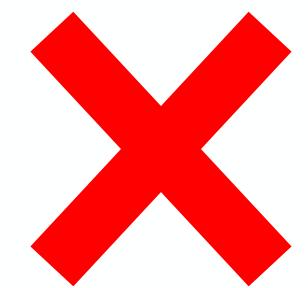
model	Precision	Recall	F1
baseline	0.600765306	0.965163934	0.740566038
chatgpt	0.839142091	0.641393443	0.727061556
chatgpt self-consistency	0.849462366	0.647540984	0.734883721
gpt-4	0.763605442	0.920081967	0.834572491
gpt-4 self-consistency	0.766609881	0.922131148	0.837209302
gpt-4-en	0.741496599	0.898969072	0.812674744
llama13b-2k	0.67614534	0.87704918	0.763603925
llama13b-2k-zh	0.675895765	0.850409836	0.753176044
llama-13b-2k-zh-self-consistency	0.700325733	0.881147541	0.780399274
llama30b-2k	0.717741935	0.911885246	0.803249097
llama65b-2k	0.73830156	0.878350515	0.802259887
llama-30b-4k-1	0.77238806	0.848360656	0.80859375
llama-30b-4k-self-consistency	0.772401434	0.883196721	0.824091778
llama-65b-4k	0.75042735	0.905154639	0.820560748



【文章】生育无限制—马来西亚不仅没有计划生育规定，而且政府鼓励多生、优生。想生就生，不受限制。拿到红卡后新生儿将自动获得大马国籍。

【谣言点】在马来西亚出生的孩子就可拥有马来西亚护照或国籍  
判断依据：

1. 【文章】提到拿到马来西亚红卡后新生儿将自动获得大马国籍。
2. 【句子】提到在马来西亚出生的孩子就可拥有马来西亚护照或国籍。
3. 拥有马来西亚国籍意味着可以拥有马来西亚护照。
4. 【文章】可以推出【句子】，是。



【文章】生育无限制—马来西亚不仅没有计划生育规定，而且政府鼓励多生、优生。想生就生，不受限制。拿到红卡后新生儿将自动获得大马国籍。

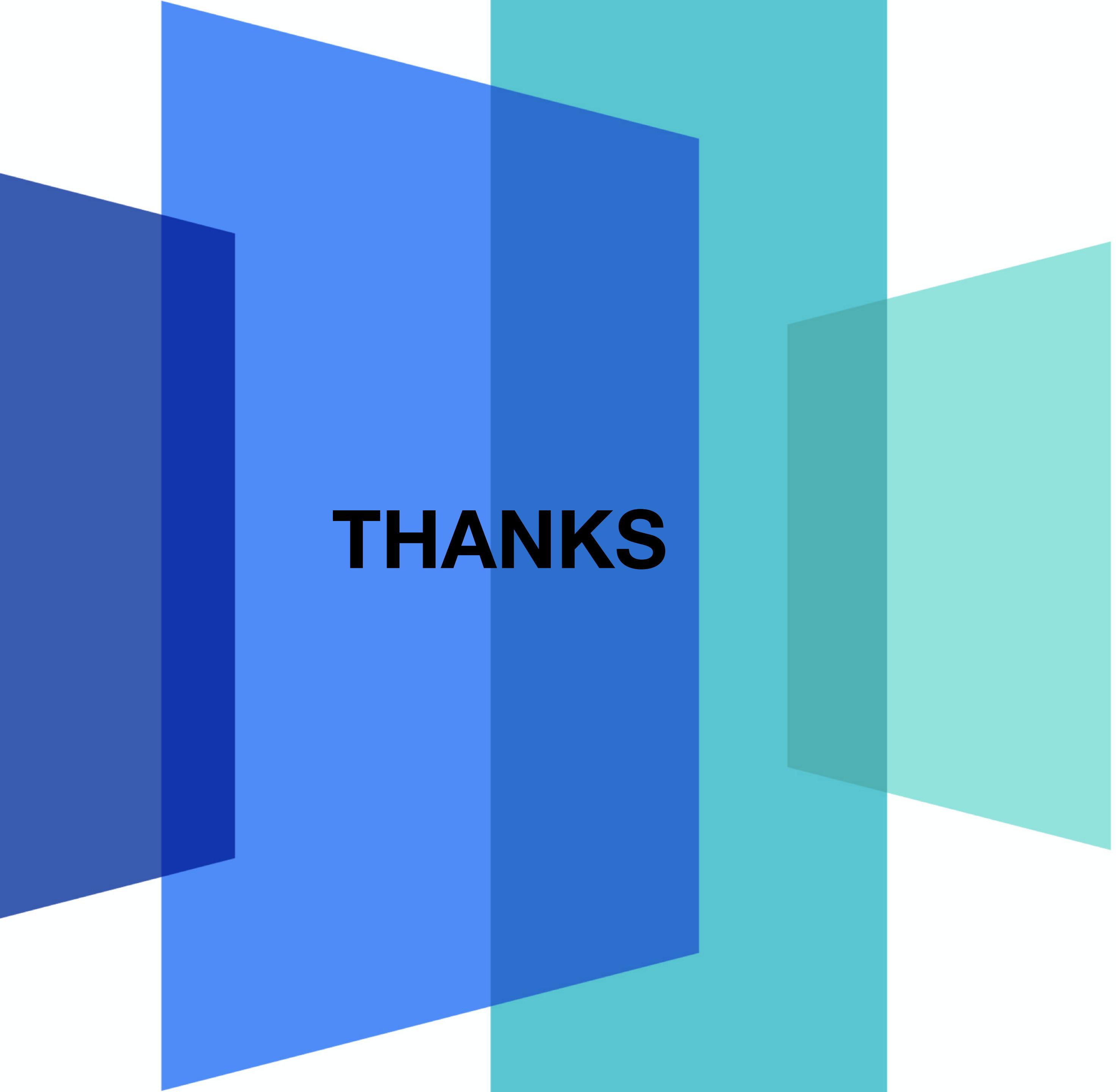
【谣言点】在马来西亚出生的孩子就可拥有马来西亚护照或国籍  
判断依据：

1. 【文章】提到拿到马来西亚红卡后新生儿将自动获得大马国籍。
2. 【句子】提到在马来西亚出生的孩子就可拥有马来西亚护照或国籍。
3. 不是所有新生儿都能拿到马来西亚国籍，只有父母拿到红卡后，新生儿才能拥有马来西亚国籍。
4. 【文章】不能推出【句子】，否。

【文章】生育无限制—马来西亚不仅没有计划生育规定，而且政府鼓励多生、优生。想生就生，不受限制。所有新生儿将自动获得大马国籍。

【谣言点】在马来西亚出生的孩子就可拥有马来西亚护照或国籍  
判断依据：

1. 【文章】提到拿到马来西亚红卡后新生儿将自动获得大马国籍。
2. 【句子】提到在马来西亚出生的孩子就可拥有马来西亚护照或国籍。
3. 拥有马来西亚国籍意味着可以拥有马来西亚护照。
4. 【文章】可以推出【句子】，是。



THANKS



XXXXXX



Xxxxxxxxxxx

# 健康专审 文本二分类

<https://bytedance.feishu.cn/docx/S7QedwGytolCOmxBe7PcrZMJnfh>

# 实体链接

- 现有业务的模型是一个在多个任务上finetune好的预训练模型，主要的任务包括抽取短答案，摘要飘红等任务。
- 当前模型的输入为一个query以及一个doc, doc包括待查询文章的title和content, content是短句构成的列表，将[CLS]+query+[SEP]+title+[CLS]+short\_sentence1+[CLS]+short\_sentence2+...作为模型输入，模型的输出再经过一系列head，最终达到业务效果。
- 目前希望在原有业务不变的基础上，加入一个实体链接的任务，将抽取出的短答案(mention)和实体库中的实体进行配对。比如：  
"query": "十部公认最好看日本电影"  
"title": "岛国最难以磨灭的10部电影佳作"  
"doc": "岛国作为亚洲电影的先锋，无论从起步还是影响力都是醒目的存在。二战之前，日本的电影工业已经非常完备，战后则开始大师涌现，诞生无数杰作。《罗生门》黑泽明撒谎是人之本性，在大多数时间里我们甚至都不能对自己诚实，但那是因为人们太脆弱了，所以才撒谎，甚至对自己撒谎。我甚至听说过在罗生门这儿的鬼，因为害怕人类的凶残而逃走。《情书》岩井俊二“当岁月流逝，所有的东西都消失殆尽的时候，唯有空中飘荡的气味还恋恋不散，让往事历历在目。”《无人知晓》是枝裕和没有哭喊没有大笑没有愤怒，却有着刺骨的冰冷...这是人世的炎凉与无奈。无人知晓这四个孩子的生活的窘境，就像无人知晓那粉色行李箱中的尸体一样...《入殓师》泷田洋二郎向死而生，学会告别。死亡不过是一扇门，让我们走向一段崭新的旅途。当逝者安息，当生者坚强，当释怀的爱留在心里，片刻即永恒，万物也许就这样生生不息..."

entity\_info: [

" 入殓师 书籍 2010年日本百濑忍所著小说 电影《入殓师》的同名小说《入殓师》在日本由著名的出版社小学馆出版。中文小说《入殓师》由北京的东方出版社引进，并于2010年4月出版。 形影,东方出版社",

" 入殓师 职业 为死者还原未死之状态的人员 入殓师又叫做葬仪师，为死者还原未死之状态。整修面容和身体，尽可能还原完整面容和身体。也可叫做为死者化妆整仪，纳入棺中的职业。主要出现在日本，后进入中国。让已经冰冷的人重焕生机，给她永恒的美丽；还要有冷静准确，并且怀着温柔的情感。 葬仪师,用杀菌皂清洗尸体,departures",

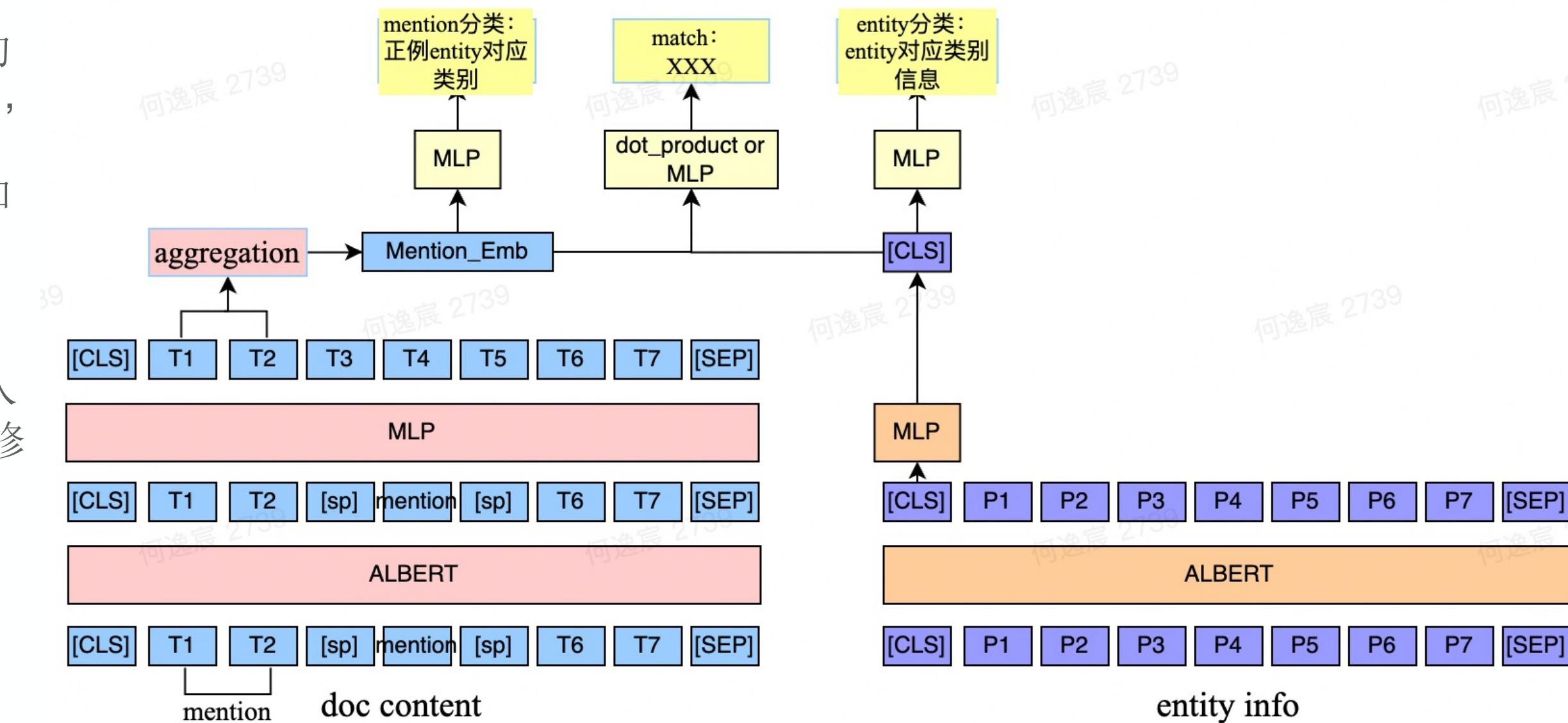
" 入殓师 电影 泷田洋二郎执导电影 《入殓师》是根据日本作家青木新门的小说《门纳棺夫日记》改编而成，由泷田洋二郎执导，本木雅弘、山崎努、广末凉子、吉行和子和 笹野高史等联袂出演。影片讲述了日本入殓师的生活，影片以一名入殓师 彩色,剧情,音乐,纪录片,为逝者送行的人,中泽敏明, 笹野高史, 浜田毅, 剧情, 山崎努, 广末凉子, Okuribito, 小越勇辉, 宫田早苗, 革命, 送行者-礼仪师的乐章, 小林美香, 日语",

" 入殓师 虚拟角色 《阴阳师》中的式神,网易手游《阴阳师》中的SR式神之一 入殓师，网易手游《阴阳师》中的SR式神之一。又称葬仪师，为逝者还原未死之状态，整修面容和身体，尽可能还原完整面容和身体的男子。 男,抽符,碎片合成,Okuribito,控制,输出,おくりびと,织雪,葬仪师,抽符、碎片合成,速水奖",

" 入殓师 小说 2016年刘东西创作的网络小说 我是一名入殓师，一天晚上，就在我和女同事在停尸间进行不可描述的事情之后，一具尸体莫名其妙的消失了..... rulianshi,巨匠阅读网",

## 现有做法

- 为了不影响原有模型的效果，将doc的侧的模型freeze住，不更新参数，用实体链接的训练集训练entity info模型。
- freeze住两个模型，训练出一个teacher model，和之前业务上的其他模型一起蒸馏。
- 这两种做法结束之后都需要将entity info侧的所有实体的embedding输出成tensor存在显存中，因此需要将768维的向量降维成128维，使用pca 初始化降维矩阵可以对墨香效果有很大的提高。



# 多模态

# 通用多模态审核模型

没有简单负例时模型会过度解读

17553081 23010052	2022年 11.12月份 新冠死亡 人数	14	2022年11.12月 份新冠死亡人数 2022年末大陆人 口141175万人， 死亡人口1041万 人。从以往死亡 情况看，冬天月 份，死亡率比春 夏秋（3~10月） 会高出1/4。 2022年春夏秋每 月平均死亡80万， 冬天月份平均死 亡100万。从火 葬场火化情况看， 2022年11.12月 份死亡人数会比 以往多2022年 11.1	我国 累计 新冠 死亡 人数 为 0.56 万	The number of COVID-19 deaths in November and December 2022 At the end of 2022, the population of mainland China 1.41175 billion, and the death population 10.41 million. From the perspective of past deaths, the death rate in winter months will be 1/4 higher than that in spring, summer and autumn (March to October). The average monthly death rate in spring, summer and autumn 2022 800,000, and the average death rate in winter months 1 million. From the perspective of cremation in crematoria, the number of deaths in November and December 2022 will be more than before. 2022 11.1	Cu mul ativ e CO VID -19 dea ths in my cou ntry 5,60 0	[Premise] discusses the possibility of a higher death rate in the winter months and the potential impact on the number of COVID-19 deaths in November and December. It is pointed out that the average monthly death rate in spring, summer, and autumn in 2022 was 800,000, and the average death rate in winter months is expected to be higher. However, it is also mentioned that the number of deaths in November and December will still be higher than usual due to the limited capacity of crematoria. These discussions suggest that the [Premise] entails the [Hypothesis] that the cumulative number of COVID-19 deaths in the country will exceed 5,600. As the [Premise] explicitly mentions the cumulative number of COVID-19 deaths and projected numbers for November and December, it can be concluded that the [Premise] does entail the [Hypothesis]. Answer: Yes.	TRUE	0	过度解读
----------------------	--------------------------------	----	--	--	---	--	---	------	---	------

1758256624 952327	43	它具人滋阴 养肝、能够 固肾、对血 压的调节也 有一定的作 用。鲍鱼有 养阴的功效， 多吃也没有 害处。	清蒸鲍鱼的简单做法，美 味佳肴，一学就会。鲍鱼特别适合女性朋友吃，它具 人滋阴养肝、能够固肾、 对血压的调节也有一定的 作用。鲍鱼有养阴的功效， 多吃也没有害处。鲍鱼的 肉中含有一种叫“鲍素”的 物质，可以起到抗癌的作 用。鲍鱼算一种比较名贵 的海产品吧。价格不菲。 今年过年我做年夜饭，做 鲍鱼是这样做的，清蒸鲍 鱼：把鲍鱼肉从壳中剥出 来，把肠肚去除掉、细沙 洗净，再用盐水清洗鲍鱼 肉。鲍鱼的壳也要清洗 干净，用刷子	鲍鱼可 以降血 压	The simple method of steaming abalone, delicious food, abalone is especially suitable for female friends to eat as soon as you learn it. It nourishes the yin and nourishes the liver, can fix the kidneys, and also has a certain effect on the regulation of blood pressure. Abalone has the effect of nourishing the yin, and it is not harmful to eat more. The meat of abalone contains a substance called "abalone", which can play an anti-cancer role. Abalone is a relatively expensive seafood. The price is expensive. This year, I made New Year's Eve dinner. I made abalone like this. Steamed abalone: peel the abalone meat out of the shell, remove the intestines, wash with fine sand, and then wash the abalone meat with salt water. The shell of the abalone should also be cleaned with a brush.	Abalon e can lower blood pressur e	[Premise] introduces a simple method for steaming abalone, which is a nutritious food that nourishes the liver and kidneys, and can also fix imbalances in the yin and yang. The meat of abalone contains a substance called "abalone" that can have anti-cancer effects, and the price of abalone is relatively expensive. For New Year's Eve dinner, the author made abalone by the method presented in the [Premise]. The steps include peeling the meat out of the shell, removing the intestines, washing with fine sand, and washing the meat with salt water. The shell should also be cleaned with a brush. In another place in the [Premise], it is also mentioned that abalone can lower blood pressure. Therefore, [Premise] can entail [Hypothesis]. Answer: Yes.	TRUE	0	过度解读
----------------------	----	--	---	-----------------	---	---	--	------	---	------

# 通用多模态审核模型

构造例子时的细微尺度差异，告诉模型尺度

张雨辰

欢迎试用AI-Lab孵化的英语写作助手——火山写作: <https://writingo.net/>

11:32

昨天看知识库bad case的时候，我们能不能这样把bad case加工成训练数据。比如s1 = "跷二郎腿、久坐... 可能会增加血栓的概率" s2 = "只要跷二郎腿就会血栓"。模型认为是蕴含，真实label是不能蕴含。我们构造两条训练数据。第一条训练数据是修正模型的输出（推理步骤和最终结果）获得，例如在输出中加入“[hypothesis]认为跷二郎腿必定导致血栓，而[premise]只说了可能增加血栓的概率，并没有必定的因果关系”。第二条训练数据是保留模型的输出，但修改s1或s2获得，比如把s2修改为“跷二郎腿和血栓有正相关的关系”。这两条样本放在一起，可以让模型学到：推理的某一步对齐了输入的某个差别，而且让它能把握模糊推理的度？

冯沛源林苑

2 条回复

新消息

林苑

11:35

回复 张雨辰：

昨天看知识库bad case的时候，我们能不能这样把bad case加工成训练数据。比如s1 = "跷二郎腿、久坐... 可能会增加血栓的概率" s2 = "只要跷二郎腿就会血栓"。模型认为是蕴含，真实label是不能蕴含。我们构造两条训练数据。第一条训练数据是修正模型的输出（推理步骤和最终结果）获得，例如在输出中加入“[hypothesis]认为跷二郎腿必定导致血栓，而[premise]只说了可能增加血栓的概率，并没有必定的因果关系”。第二条训练数据是保留模型的输出，但修改s1或s2获得，比如把s2修改为“跷二郎腿和血栓有正相关的关系”。这两条样本放在一起，可以让模型学到：推理的某一步对齐了输入的某个差别，而且让它能把握模糊推理的度？

@张雨辰 好的，现在是拿GPT-4的数据直接训练，后面把人review的流程加进去，如果做错了，就按照这个方法同时构造两条正确数据。

张雨辰

张雨辰

欢迎试用AI-Lab孵化的英语写作助手——火山写作: <https://writingo.net/>

11:37

回复 林苑：

@张雨辰 好的，现在是拿GPT-4的数据直接训练，后面把人review的流程加进去，如果做错了，就按照这个方法同时构造两条正确数据。

@林苑 这种构造方式可以确保两条样本的s1和s2很相似，但最终的结论是相反的（必定是一个“是”一个“否”）。这一点比较重要，因为是在帮模型在“是”与“否”之间确定一个边界。