

SpY第七次会议记录 5.8

参会人：袁晨 王柏霏 尹天禹 张瑞林 杜周纬 赵培炎

会议时间：5月8日 20:00-21:00

会议地点：腾讯会议

迭代三讨论

- 自动化情报收集，接入Flink等流处理工具对新闻爬取到的新闻进行实时处理，实现自动化情报收集、分析及展示
 - 接入流处理工具，以实现可以实时抓取新闻页面上的更新并进行分类和数据分析功能。（链接：<https://news.sina.com.cn/roll/>）
 - 前端页面需要随着新闻的更新而动态实时变更，如哪个分类新增了哪条新闻等。
 - 爬虫写推送，rabbitmq，flink取，存之前需要处理，需要存三份
 - 两小时爬1~2条，爬几天就要停掉
 - 前端不太需要实时更新
- 情报可信性分析，可尝试加入时间轴分析，新闻虚假鉴定等模块，分析情报的可信度。
 - 加入新闻可信性分析，判断该条新闻的真实指数。
 - 实现时间轴分析
 - 可以搞一下，效果不保证
 - 不懂，需要研究什么是可信
 - 尝试使用大模型
- 大模型情报分析，接入大语言模型，实现实时问答式情报分析，处理类似于“当日体育类新闻的条数有多少？”，“相比于昨日，时政类新闻同比增长多少？”等问题。
 - （加分点）接入大语言模型API（包括但不限于GPT等），以实现大语言模型新闻情报分析功能。
- 尝试提高迭代2中知识图谱生成的准确率。
 - （加分点）提高知识图谱生成的准确率
 - 记录下来
- 完善并整合迭代1、迭代2及迭代3，最终实现一个完整的新闻情报分析系统产品。

1. 模型训练
 - a. 问一下
2. 前后端开发，
 - a. 主分支是后端master，前端master
 - b. 以后开发功能，自己新建一个分支，写完，合并到主分支里，yuanchen_homeModify

任务分工

- ☐ CICD 需要买一个服务器，前端部署修改 王柏霏
- ☐ Prompt 亮点设计 编写 杜周纬 待定
- ☐ 知识图谱准确度提升策略 编写 杜周纬 待定
- ☐ 每次都提交上去会议记录 袁晨
- ☐ Flink
 - ☐ 搭环境 尹天禹 赵培炎辅助
 - ☐ 存三份 赵培炎 neo4j 问张瑞林
 - ☐ 爬虫 rabbitmq 推送 杜周纬
- ☐ 可信性分析 王柏霏 袁晨调研 张瑞林 main
 - https://www.researchgate.net/publication/335082931_Predicting_News_Source_Credibility 根据新闻来源分析
 - 微博可信性<https://ieeexplore.ieee.org/document/7023340>
 - 根据作者判断假新闻 <https://ieeexplore.ieee.org/document/7023340>
 - <https://arxiv.org/pdf/2402.10938> 神经网络
 - 需要寻找根据内容特征判断假新闻的途径
 - 关键词：content features、News credibility analysis model
- 1. 依据数学建模的一些综合评价方法，如优劣解距离法、层次分析法等
 - a. 层次分析法求权重
 - i. 确定层次结构：首先确定评估层次结构，包括目标层、准则层和方案层。在这里，目标是评估新闻的可信度，准则包括来源权威性、语言合理性、逻辑合理性和情感倾向。
 - ii. 构建判断矩阵：针对每个准则，构建一个判断矩阵，用来比较不同准则之间的相对重要性。例如，你可以让专家根据其经验和直觉，给出不同准则之间的两两比较，然后将这些比较转化为数值。
 - iii. 计算权重：使用AHP方法计算每个准则的权重，这可以通过对判断矩阵进行特征向量归一化来实现。

- iv. 评估方案：对于每个新闻报道，根据各个准则的权重以及相应的子准则，评估其可信度得分。
- v. 综合评估：将各个准则的评估结果综合起来，得到最终的新闻可信度评分。

可信度	来源	语言	逻辑	情感
来源	1	2	2	4
语言	1/2	1	1	3
逻辑	1/2	1	1	3
情感	1/4	1/3	1/3	1

b. TOPSIS法（优劣解距离法）具体打分

可信度	来源	语言	逻辑	情感
来源	1	2	2	4
语言	1/2	1	1	3
逻辑	1/2	1	1	3
情感	1/4	1/3	1/3	1

将判断矩阵按照列归一化

可信度	来源	语言	逻辑	情感
来源	0.4444	0.4615	0.4615	0.3636
语言	0.2222	0.2308	0.2308	0.1818
逻辑	0.2222	0.2308	0.2308	0.1818
情感	0.1111	0.0769	0.0769	0.0909
	2.25	13/3	13/3	11

得到权重向量

--	--

	计算出可信度权重
来源	0.43275
语言	0.2164
逻辑	0.2164
情感	0.13445

2. **新闻来源可信度评估**：评估新闻来源是否权威官方。根据权威机构和官方榜单来确定权威的媒体名单。如果来源不在确定的权威媒体名单内，可能需要对其可信度进行怀疑。

a. 媒体分类

i. **财经、股票、经济、房地产类媒体**：

1. 21世纪经济报道、第一财经、每日经济新闻、财新、财经杂志、经济参考报、澎湃新闻、虎嗅、36Kr、蓝鲸财经、界面新闻、钛媒体、证券时报、证券日报、中国证券报、上海证券报、财经网、观察者网
2. 中国证券报
3. 证券时报
4. 上海证券报
5. 证券日报
6. 财经网
7. 东方财富网
8. 中证网
9. 华尔街见闻
10. 金融界
11. 同花顺财经
12. 新浪财经
13. 雪球财经
14. 东方财富
15. 21世纪经济报道
16. 中国金融信息网
17. 经济学人
18. 新浪、凤凰、网易、搜狐、腾讯
19. 财经 财联社 商界 财新 创业邦

20. 第一财经 虎嗅 投资界
21. 界面新闻 华商韬略 证券日报
22. 经济观察报 华夏时报 中国基金报
23. 经济日报 蓝鲸财经 中国经济周刊每日经济新闻 新财富 中国企业家
24. 央视财经 证券时报&券商中国 中国证券报
25. 中新经纬 中国经营报 中欧商业评论
26. Vista看天下 观察者网 环球时报 环球网 澎湃新闻 人民日报 新华社
27. 新京报 贝壳财经 中国青年报 中国新闻社
28. 补充

证券时报

全景网络

和讯中金网

中国企业家

中国证券网

中国证券报

新浪网财经纵横

财经

财富时报

长江证券

江苏经济报

经济参考报

环球企业家

中国企业报

21世纪经济报道

赛迪网

成功营销

期货日报

中国经营报

粮油市场报

中国食品网

天下人才网

国际经贸信息网

新财经

证券导报

环球财经

海峡消费报

河北经济日报

中国经济时报

中国科技财富

新华信管理咨询

新元网

西南商报

深圳商报

中国金融网

中国房地产报

河北经济日报

市场报

信息早报

全球纺织网

商务部网站

北京现代商报

黑龙江经济报

贵州商报

财经时报

中国商贸

四川经济日报

粤港信息日报

企业文明杂志

期货日报（河南）

现代物流报

今传媒

金融时报

ii. 时政、社会类媒体

1. 新华社，参考消息，人民日报，紫光阁，半月谈、澎湃
2. 参考消息、人民日报、环球时报、南方周末、光明日报、广州日报、南方都市报、中国青年报、经济日报
3. 腾讯、新浪、搜狐、网易、凤凰、新华、人民
4. 中央人民广播电台、中国国际广播电台
5. 中央电视台、浙江卫视、湖南卫视、东方卫视、江苏卫视、北京卫视
6. 人民网、新华网、求是理论网、中国军网、光明网、中国经济网、央广网、央视网、国际在线、中国日报网、中国科技网、中国纪检监察报社网站、中工网、中国青年网、中青在线、中国妇女网、中国农业新闻网、法制网、中国新闻网、中国网、新华社、中国消费者协会、中国广播网
7. 中央新闻、央视网
8. 人民日报
9. 新华社
10. 中国中央电视台（CCTV）
11. 中国经济时报
12. 中国新闻社
13. 中国青年报
14. 光明日报
15. 经济日报
16. 中国日报
17. 中国政府网
18. 中国新闻网
19. 中国国际广播电台
20. 中国广播网
21. 中国青年网
22. 中国经济网
23. 中国共产党新闻网
24. 中国广播电视台
25. 中国广播电视网
26. 中国国际广播网

27. 中国军网
28. 中国法制网
29. 中国人民解放军报
30. 中国纪检监察报
31. 中国消费者报
32. 中国科技日报
33. 中国社会科学报
34. 中国教育报
35. 中国新闻周刊
36. 中国新闻出版报
37. 中国外交部网站

iii. 科技类媒体

- 品玩
 - 爱范儿
 - 36 氪
 - 虎嗅
 - 极客公园
 - 奇客Solidot
 - 少数派
 - IT之家
 - 钛媒体
 - 量子位
 - 机器之心
 - 雷锋网
1. 中国科学报
 2. 科技日报
 3. 中国科技网
 4. 科技导报
 5. 中国科技信息
 6. 科技风
 7. 科技日报社

8. 亿欧网
9. 电脑报
10. 环球科技

iv. 家居类

1. 新浪, 腾讯, 网易, 搜狐, 凤凰
2. 太平洋家居网, 一兜糖, 好好住, 住范儿、今日家居、家居头条
3. 瑞丽家居, 时尚家居, ELLE家居廊、
4. 家居观察
5. 家居资讯网
6. 家居科技
7. 家居传媒
8. 家居博览
9. 家居商业评论
10. 家居行业网
11. 家居时代

v. 教育类

1. 芥末堆、多知网、鲸媒体、36kr、澎湃、中国
2. 腾讯、新浪、搜狐、网易、凤凰、新华、人民
3. 中国教育新闻网
4. 教育在线
5. 教育部网站
6. 中国教育报

vi. 时尚类

1. YOKA、NOWRE、瑞丽、时尚芭莎、GQ、嘉人、VOGUE、时装L'OFFICIEL、ELLE、COAMO、CFW、FHM、GRAZIA、时尚头条、LADYMAX
- 凤凰网时尚、新浪网时尚、环球网时尚、北青网时尚、TOM时尚、国际在线时尚、新华网时尚、搜狐网时尚、太平洋时尚网、人民网时尚

vii. 体育类

1. 中国体育报、体坛周报、足球周刊
2. 体育报
3. 体育时报

4. 体育频道
5. 体育文化报
6. 新浪
7. 腾讯
8. 网易
9. 搜狐
10. 人民
11. 凤凰
12. 央视网
13. 虎扑
14. CCTV
15. 中国体育报
16. 中国体育电视台
17. 中国新闻社体育部

viii. 游戏类

1. 游戏葡萄, Gamelook, 游戏陀螺
2. GameRes、Indienova、老东家
3. 玩加、PentaQ、cowlevel
4. 36Kr, Pingwest、NEXT Studios
5. 篝火营地、taptap
6. 游研社、触乐、机核、游资
7. Tgbus, 多玩, 游民星空
8. 游戏机实用技术、大众软件
9. 新浪、网易、腾讯、搜狐

ix. 文娱类

1. 新浪、网易、腾讯、搜狐
2. 中国娱乐网、半岛、央视、凤凰
3. 美术报
4. 中国消费网
5. 新浪网生活空间
6. 乐视

- 7. 中华文摘
- 8. 中华读书报
- 9. 中华集邮网
- 10. 北影画报
- 11. 中国文化报
- 12. 精品购物指南
- 13. 中国艺术报
- 14. 家庭导报

x. 星座类

- 1. 默认分数

xi. 中央媒体（关键字样：人民、中国、央、日报、新华、

人民日报

新华社

中国青年报

国际广播电台

人民邮电报

央视国际

国际在线

华夏经纬网

中国质量万里行

光明日报

中国台湾网

世界新闻报

中国质量报

中国新闻周刊

中国互联网新闻中心

人民公安报

工人日报

中新社

中国教育报

环球时报

中国记协

中国纪检监察报

检察日报

中国信息报

解放军画报

人民法院报

中国人事报

中国医学论坛报

中国日报

半月谈

中国交通报

中国气象报

中国化工报

中国绿色时报

小康

外交部网站

华南新闻

中国人口报

旅游中国周刊

法制日报

中国妇女报

科技日报

中国国防报

求是

商务部网站

华东新闻

经济日报

人民政协报

人民铁道报

中国涉外商事海事审判网

中国经济网

中国摄影报

解放军报

中国广播网

现代教育报

三农在线

国际商报

农民日报

绿色中国网

中国旅游报

中国建设报

中国消费者报

健康报

中国审计报

中华建筑报

中国食品质量报

中国产经新闻报

健康时报

劳动午报

中国食品报

中青传媒

中国远洋报

中国邮政报

中国渔业报

中国新闻出版网

中国海洋报

中国航天报

中国黄金报

中国电力新闻网

绿色中国网络电视

中国劳动保障报

中国低碳经济网


中国数字期刊行业资讯

中国民族报


中国西藏网

3. 新闻语言分析：分析新闻报道的语言是否合理。


- f. 大模型分析
- g. 提示词（已测试）：请为这篇新闻的语言合理性进行打分，只需要分数，满分为100分，输出数字即可，要敢于打低分和高分



美国总统乔·拜登（Joe Biden）周三上午抵达沃尔特·里德国家军事医疗中心（Walter Reed National Military Medical Center）进行年度体检，预计将在下午晚些时候正式公布他的体检总结报告。虽然此次体检是一项例行检查，但今年它对这位81岁的现任总统来说事关重大，因为许多民众怀疑拜登的身体和精神能力能否支持他在白宫再任职四年。这位民主党人经受住了政治对手对他的年龄和健康状况的无情攻击，尤其是共和党总统候选人唐纳德·特朗普，后者将于今年6月年满78岁。在2023年的体检之后，白宫医生凯文·奥康纳（Kevin O'Connor）博士表示，拜登“仍然是一个健康、精力充沛的80岁男性，他适合成功履行总统的职责”。请为这篇新闻的语言合理性进行打分，只需要分数，满分为100分，要敢于打低分和高分，回答输出数字即可



92分。这篇新闻的语言表达清晰，信息准确，没有出现明显的语病或逻辑错误。同时，它也成功地传达了新闻的主要内容，即美国总统乔·拜登在沃尔特·里德国家军事医疗中心进行年度体检，并在稍后公布体检报告。

 重新回答

由星火AI生成

当地时间9日晚，以色列国防军发言人丹尼尔·哈加里在新闻发布会上表示，以军正在加强斋月前的安全准备，已准备好“应对各种可能情况”。哈加里还称，巴勒斯坦伊斯兰抵抗运动（哈马斯）拒绝了新一轮加沙停火协议，拒绝接受停火谈判中各斡旋方的建议。他说，以色列情报显示，哈马斯正在为其武装人员囤积食物和装备。哈马斯方面对此暂无回应。分析这篇新闻报道的语言是否合理。请为这篇新闻的语言合理性进行打分


这篇新闻的语言合理性打分：8/10

由星火AI生成


- i. 准确性：新闻语言应该准确地传达事实，避免虚假信息或误导性陈述。关键信息应该经过验证和确认。
 - ii. 客观性：新闻语言应该客观中立，避免主观偏见或情绪化的表达。使用客观的措辞和描述来呈现事件或主题。
 - iii. 清晰度：新闻语言应该清晰易懂，避免使用模糊不清或含糊的词语，避免使用过度复杂的句子结构。
 - iv. 公正性：新闻语言应该公正公平，尊重各方立场和观点，避免歧视性言论或偏袒某一方。
 - v. 完整性：新闻语言应该全面报道事件的各个方面，避免片面或不完整的报道，确保读者能够全面了解事件。
4. 新闻逻辑分析：分析新闻逻辑是否合理。如果新闻报道中存在明显的逻辑漏洞，可能需要对其可信度进行怀疑。

a. 大模型分析

- b. 提示词（已测试）：请为这篇文章的逻辑进行打分，只需要分数，满分为100分，输出数字即可，要敢于打低分和高分







我吃饭了，但我又饿了。他在打球，同时在写作业。请为这篇文章的逻辑进行打分，只需要分数，满分为100分，输出数字即可，要敢于打低分和高分




40分

由星火AI生成



术。尽管许多UAP报告仍未得到解决或身份不明，但AARO评估称，如果有更多、更优质的数据可用，这些案例中的大多数也可以被识别和解决为普通物体或现象。自1945年以来，美国政府一直在资助调查，以确定UFO是飞行安全风险、竞争国家的技术飞跃，抑或是高智能外星技术。几十年来，世界都在盛传美国政府内部的一个秘密组织已经发现了几艘“外星飞船和外星生物遗骸”，并对这些发现展开了“逆向工程”项目。AARO指出，许多人都相信这些传言，甚至成为了流行文化，但工作人员使用了严格的分析和科学的方法，调查了过去的UAP报告，结果可能会让人们失望，没有发现任何证据表明美国政府和私营公司一直在对外星技术进行逆向工程。五角大楼两年前表示，他们的调查工作带来了数百份新的报告，但没有任何证据表明存在外星智慧生命。与此同时，美国军方正在开发一种便携式不明飞行物探测工具，以收集更准确的目击报告数据。据报道，正在开发的便携式探测工具由一系列传感器组成，目前正在得克萨斯州进行测试。请为这篇文章的逻辑进行打分，只需要分数，满分为100分，输出数字即可，要敢于打低分和高分



90分

由星火AI生成

5. 新闻情感分析：新闻报道的情感倾向可以反映报道者的态度和立场。一些偏颇或不客观的报道可能会在语气和情感上表现出来，情感分析可以帮助识别这种倾向性。

- a. 过于激进的、充满煽动性的情感会降低新闻的可信度

b. Hanlp

情感分析（Sentiment Analysis）任务的目标是判断一段文本的情感极性。情感极性为 $[-1, 1]$ 之间的数值，数值的正负代表正负面情绪，数值的绝对值代表情感的强烈程度。

☐ 前端把可信性分析加上去 袁晨

☐ 问答式 袁晨 待定

☐ 计划文档 待定

☐ 需求文档 待定

截止日期 5月20号