



上海图书馆开放数据竞赛

- 2 0 2 4

# 上海万象

*Shanghai Mosaic* »

复旦大学 SoaringLab实验室  
汇报人： 蒋黎明



复旦大学  
怀  
北京大学  
北京邮电大学  
浙江大学

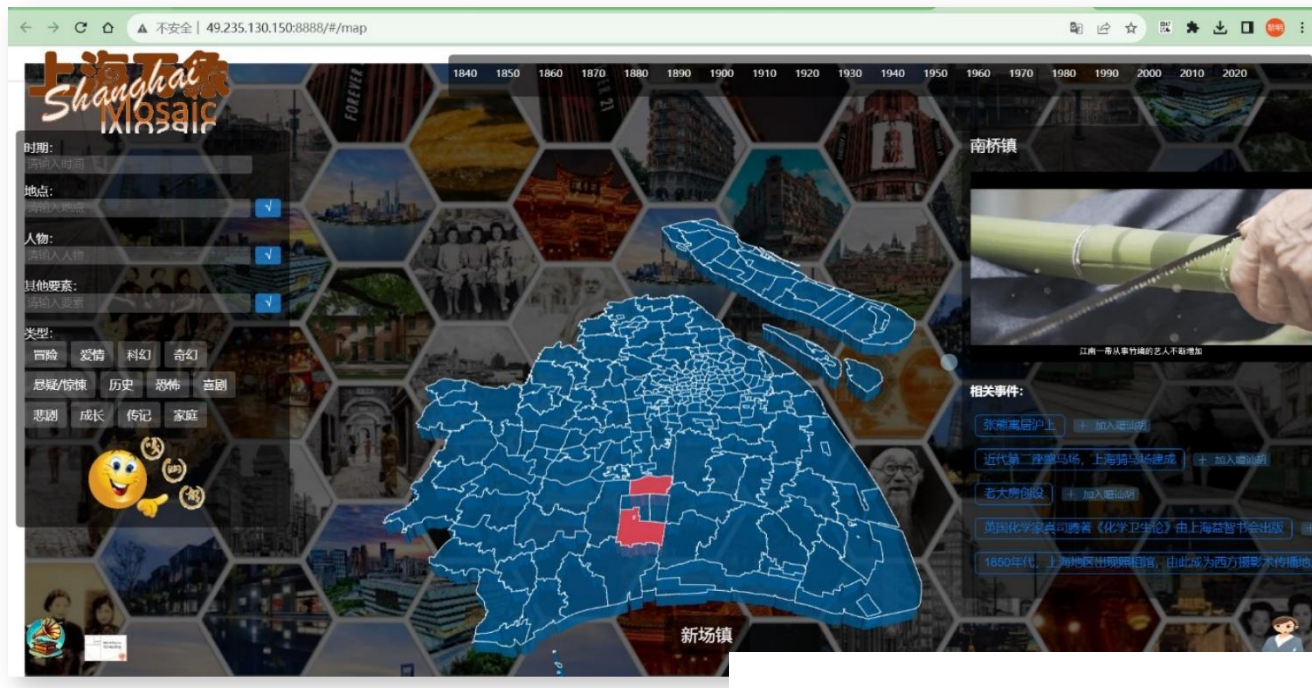
蒋黎明、朱允  
符炫炜  
高梦洁  
刘九龙

# 作品概述

Work description

“上海万象”是一款基于**多模态AI大模型+RAG技术**与丰富数据资源的创新项目，旨在全方位、多层次地呈现从十八世纪至二十一世纪初的上海风貌。该项目集成了多家著名机构的数据资源，包括上海图书馆、韬奋纪念馆、交大图书馆、虹口图书馆等，结合上海专题片、家谱等多源异构数据，构建了一个立体化、时空交互的数字上海。

项目地址：<http://49.235.130.150:8888/#/map>



上海万象

以通义千问大模型为基座，配合RAG技术，做智能化的人文历史信息检索、“噶讪胡”了解历史故事，智能推荐“专题片”，从中体会图书馆馆藏资料中蕴含的人文精神。

01

# 作品概述

时期: 民国初年时期

地点: 请输入地点  ✓

三林烈士陵园

人物: 请输入人物  ✓

其他要素: 请输入要素  ✓

扬剧女演员黄秀花生

类型:

冒险 爱情 科幻 奇幻

悬疑/惊悚 历史 恐怖 喜剧

悲剧 成长 传记 家庭

点击



## 时空

通过时间轴和上海地图双重交互界面，用户可以在特定的时空维度下，直观地浏览某个历史时期的上海城市景观、建筑、重大事件以及关键人物。

## 检索

在数据检索方面，“上海万象”超越了传统的关键词检索，采用了深度语义搜索 (Semantic Search) 与时空语义嵌入技术 (Spatio-Temporal Semantic Embedding)，能够更精准地理解用户意图，提供更贴近需求的检索结果。

## 故事

这些故事可能是如梦似幻的传奇，也可能是感人至深的叙事，甚至是充满戏剧冲突的悲剧。为了增强用户体验，模型支持多模态输出，不仅能生成文本，还可以自动匹配相关的历史图片和视频片段，通过深度学习的图像生成与视频合成技术，实现视觉与听觉的多重享受。

评委问题：  
生成的故事不是真实的，假故事会误导判断力不足的用户，就算改编故事也要显示出来引用！

“噶讷胡功能”



# 解决的主要问题——RAG

Main problem solved

## 上海万象的核心——RAG技术

检索增强生成（RAG, Retrieval Augmented Generation）技术是项目的重要技术支柱之一。RAG技术结合了大语言模型（LLM）与信息检索的优势，**在生成内容之前，会首先从预先构建的知识库或文档集合中检索相关信息，并将这些信息作为生成的参考依据。**

### CASE：民国初年的南汇发生了什么故事？ *参赛时候的CASE*

关于民国初年南汇（现属于上海市浦东新区）的具体惊悚故事记录较少，这可能是由于地方历史资料的保存情况以及当时社会对于此类事件的记录方式所限。不过，在那个时代背景下的中国，由于政治动荡、军阀混战等原因，各地都流传着不少传说或民间故事，其中不乏一些带有神秘色彩的故事。

比如，在上海及周边地区，就有许多与鬼怪相关的传说。但这些故事大多缺乏可靠的历史依据，更多地体现了当时人们对于未知事物的好奇心以及在艰难环境中寻求心灵慰藉的一种方式。

如果你对特定类型的“惊悚”故事感兴趣，如灵异事件或者悬疑案件等，建议查阅相关的地方志书、民俗学研究资料，或者是寻找一些文学作品中虚构的情节来满足好奇心。需要注意的是，很多所谓的“真实”故事往往经过了口耳相传而被夸大其词，因此在听闻时最好保持批判性思维。

如果您是指具体的某一类事件，请提供更详细的信息，以便于我能给出更加准确的回答。例如，是否有关于某位名人、某个地点或是某种特殊现象的故事传闻？这样或许能找到一些更为贴近您需求的内容。

未使用RAG技术的大模型效果



讲述了一个不相识的李阿毛和小芹发现灵异的闪光原来是他们共同父辈们的革命遗物的故事

哎，朋友，让我给您讲个老早辰光上海的故事，今朝已经属于上海浦东的一部分了，南汇烈士陵园李阿毛打小在南汇长大，对这一带的每条弄堂、谣言四起。这不，就在陵园边上那块荒废的地界了许多原本想去那里偷菜的邻家小娃。那一年夏末，正值上海医学专门学校的开学季，小芹的女孩，她的祖父曾参加过反封建的革命运动，陵园给祖辈扫墓，了却心中敬意。小芹到了南汇之后，借宿于李阿毛的家中。李阿毛那片陵园外的荒地上，关于发光物件的奇闻，说那旺盛的丫头，听了这故事不但不怕，还多了份挂念。一日傍晚，太阳还未全落下地平线，两人结伴而行，带来一丝丝清凉。进入陵园，庄严肃穆，小芹又祭祀完毕，她转向那传说中古怪的地界，拉着这天渐渐暗了下来，星星开始点缀于天空。两人作怪鸣之外，便是寂静一片。正行间，李阿毛忽然向脚下——竟是一只旧铜箱半掩在土中，而那小芹好奇心起，和李阿毛合力将那铜箱子从土里信件中记录了小芹祖辈以及他们的同志们的秘密。李阿毛异常相似的年轻人格外引人注目。原来，李阿毛的祖辈也是南汇一带秘密抗队伍以出现在这里，是因为革命成功后，一部分文件一阵夜风吹过，带来的不是寒意，而是温暖人心是一个连接过去与未来的神秘桥梁。自此以后，那些所谓的恐怖传说似乎在南汇烈士陵园形式。每当夏夜星空之下，人们走过陵园周边，就是上海南汇的故事，带着悬疑与感动。



使用了RAG技术获取“上海图书馆”事件数据的大模型效果

# 解决的主要问题——RAG

Main problem solved

## 上海万象的核心——RAG技术

检索增强生成（RAG, Retrieval Augmented Generation）技术是项目的重要技术支柱之一。RAG技术结合了大语言模型（LLM）与信息检索的优势，**在生成内容之前，会首先从预先构建的知识库或文档集合中检索相关信息，并将这些信息作为生成的参考依据。**

标签：抗日战争时期 & 朱家店抗日之战纪念地点 *改进后的CASE* 改变生成内容：图片/视频/上海话音频

朱家店抗日之战是什么时候发生的？

朱家店抗日之战发生在1944年8月21日，是新四军浙东游击纵队浦东支队指挥的一场重要伏击战。在朱亚民的带领下，他们巧妙设伏，在上海奉贤（今属闵行区）朱家店地区成功袭击了由日本鬼子、伪军以及特务共约300余人组成的队伍。这场战斗不仅展现了中国军队的智慧和勇气，也为当地抗日武装斗争留下了浓重的一笔。

要再聊聊啥？比如阿拉上海还有哪些抗日英雄或者战役？

- 复旦大学红色网站上关于朱家店之战的历史记载: [详情页面](#)
- 携程网对朱家店抗战纪念馆及其周边景点的旅游介绍: [查看详情](#)

AIGC展示数据来源

[Type your message...]

发送

万有桥有哪些重要的历史背景和相关的故事，能否提供相关文字和语音介绍资料？

关于崇明区的视频，为您推荐以下四个具有特色主题的视频：

1. **扁担戏** - 探索独特的民间木偶戏艺术。视频链接: [点击这里观看扁担戏视频](#)
2. **崇明竹编技艺** - 了解精湛的手工艺及其在现代生活中的应用。视频链接: 第1个视频
3. **崇明老白酒** - 品味地方特色酒文化及其制作工艺。视频链接: 第2个视频
4. **崇明土布纺织技艺** - 深入纺织历史，体验传统纺织技术的魅力。视频链接: 第3个视频

以上专题片均为崇明地区的文化瑰宝，既富有教育意义又引人入胜，能够让您更加全面地了解崇明的风土人情与传统工艺。请选择您感兴趣的视频进行观看。

10:51 / 29:56

对于万有桥的具体历史背景及相关故事，当前提供的资料中并没有直接提及[1]。因此，我暂时无法提供关于万有桥的文字和语音介绍资料。如果有关于其他上海红色建筑的信息需求，比如抗日第八集团军司令部旧址的历史，我可以基于现有资料进行分享。对于万有桥更详细的信息，可能需要通过其他途径进一步查询。

参考资料中提到了抗日第八集团军司令部旧址的一些信息，但这是与万有桥不同的地点。如果您对此处具有重要历史意义的建筑感兴趣，我很乐意为您提供相关信息。不过，请注意，目前我的能力仅限于提供文字描述，还不支持生成语音介绍。 [1]

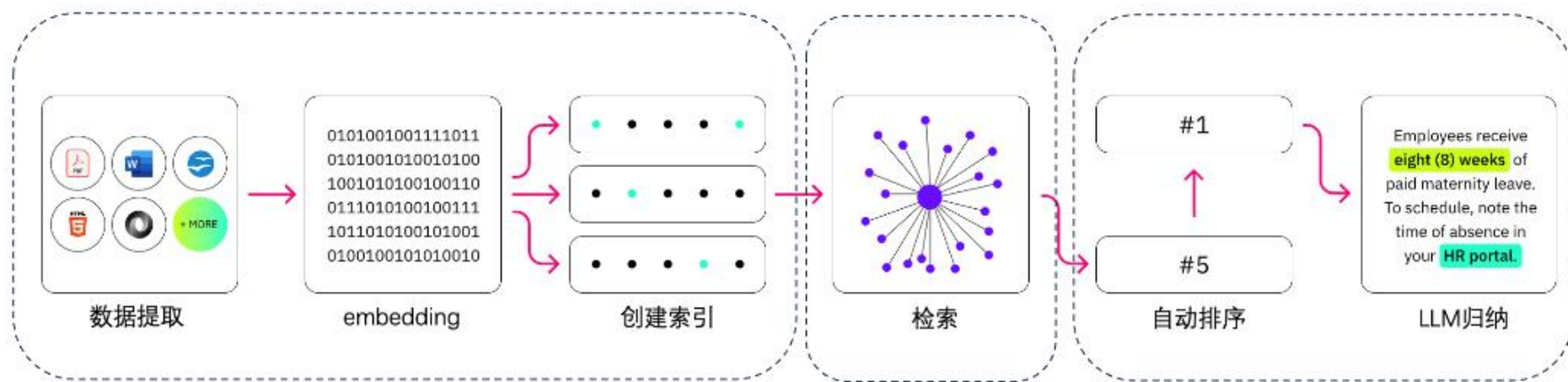
【文档名】:上海红色建筑 建筑名:抗日第八集团军司令部旧址 建筑图片链接:<https://img.library.sh.cn/shmemory/hslly/architecture/img/y83fd7ejdgtzw63a.jpg> 建筑介绍:原为南桥天主堂神火楼，1937年“八一三”淞沪抗战爆发后，至8月抗日第八集团军驻扎浦东、浦南时，邀请了郭沫若、夏行、田汉等30余人来参加奉贤县军民抗日联欢大会。11月7日第八集团军奉命撤离，南桥沦陷，该司令部撤除。

引用资料:

# 技术实现方案

Technology plan

## 大模型+RAG细节



### 数据索引

首先，将图书馆数据进行预处理，包括分块、清洗和向量化。这些数据被转化为计算机可理解的向量格式，并存储在索引中，以便于快速检索。

#### 数据索引与向量化

### 检索

#### 检索过程

当用户提出查询时，系统会根据预先建立的索引进行检索，找到与查询最相关的数据块。检索结果经过优化排序，以确保与查询的高度相关性。

### 生成

检索到的内容被输入到大语言模型中，模型会基于这些检索结果生成最终的响应文本。这个生成过程融合了检索到的信息，确保输出内容的准确性和上下文关联性。

#### 生成过程

# 技术实现方案

Technology plan

## 百炼大模型平台

大模型服务平台百炼是一站式的大模型开发及应用构建平台。可以使开发者还是业务人员深入参与大模型应用的设计和构建。提供外接RAG知识库的管理，允许管理结构化和非结构化的数据。

### 数据管理

平台统一的数据管理模块，支持多源、多格式的结构化及非结构化数据管理，支持多类目管理及自动化在线解析，应用于RAG、训练等模块。

非结构化数据    结构化数据

数据表管理 (10/1000) +

- 上海红色建筑
- 招商银行支行问答表
- 招商银行支行评论表
- 招商银行支行
- 上海视频专题片
- person2
- person
- minguo\_journal
- event2
- event

#### 上海红色建筑 导入成功

共301条数据 ↓ ☰ + 导入数据

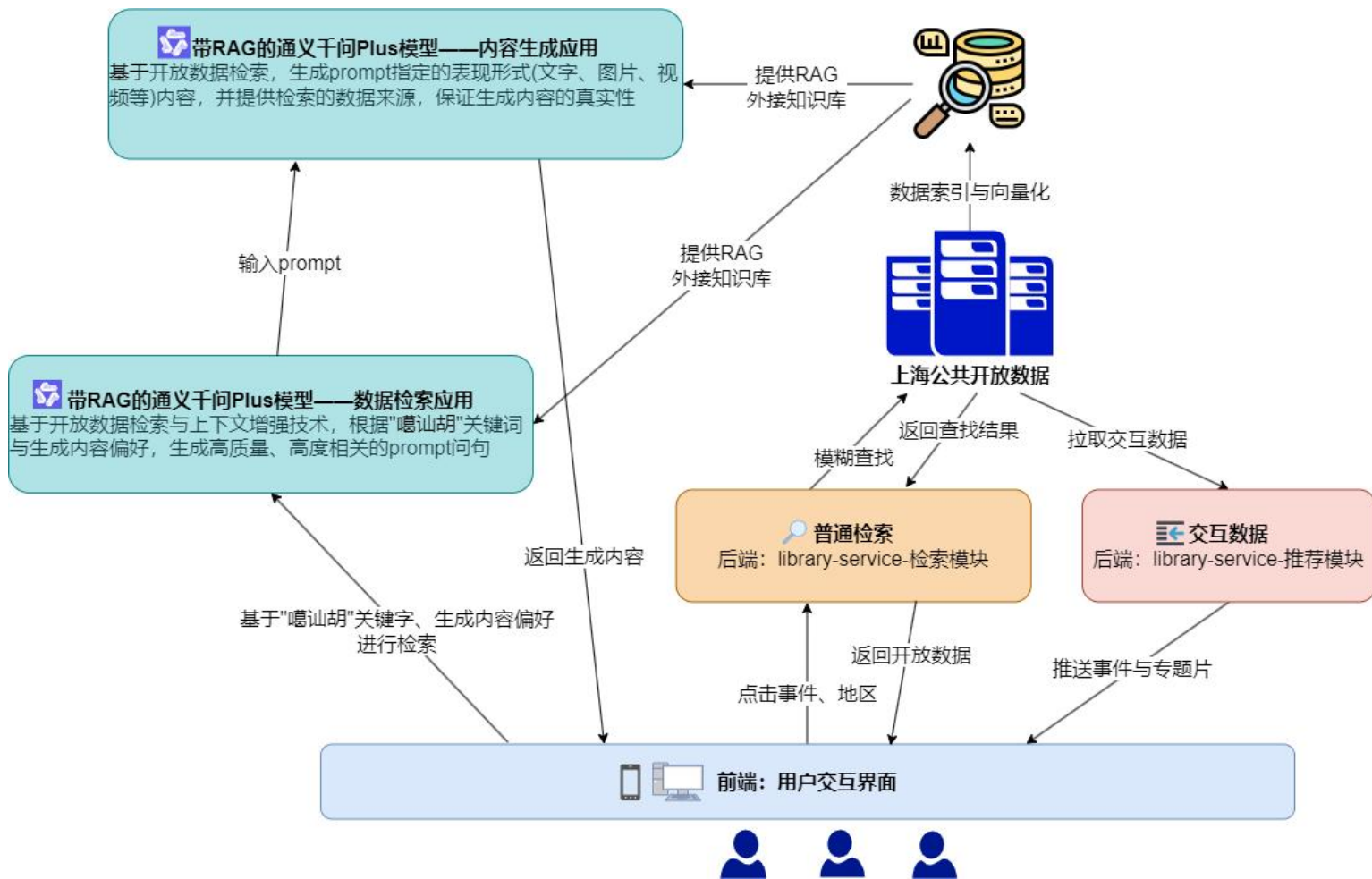
建筑名 (string)	建筑介绍 (string)
生活书店旧址	1932年7月1日成立，创办人邹韬奋。其前身是《生活》周刊社。《生活》周刊创刊于1925年10月，是中华职业教育社
大同幼稚园旧址	位于戈登路武定路口（今江宁路441号）的一幢石库门房子是1930年3月至12月大同幼稚园的旧址（原建筑已拆）。大
中国青年新闻记者协会成立大会会址	中国青年新闻记者协会成立大会旧址位于山西路194号（今山西南路182-200号）。1937年，抗日战争全面爆发，上海
刘长胜故居	刘长胜故居位于愚园路81号。该宅是一幢沿街砖木结构的三层楼房。1946年至1949年，刘长胜曾在此居住，此地也

< 1 2 3 4 5 ... 31 > 10条/页 ▾

# 技术实现方案

Technology plan

## 项目整体架构



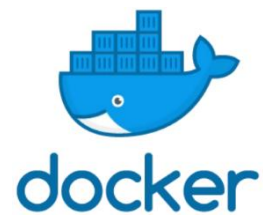
用户交互界面：Vue2.0

后端服务：SpringBoot

Docker部署在腾讯云

LLM:

不带RAG的普通大模型  
整理用户的模糊行为；  
带有上海图书馆开放数  
据的RAG知识库的大模型  
做相应回答。



# 数据使用量情况

Data used

## 上海图书馆:

上海优秀历史建筑 1086处

上海市不可移动文物 3452处

红色旅游景点 1199处

上上海市历史文化事件 1.3万余条

《近代上海城市文化》专题片 12集

海地名志 10000余个

中文期刊论文数据 1万余条

## 上海市静安区图书馆:

《静安文化专题片之非遗篇》 6集

## 上海市闵行区图书馆:

《上海闵行民俗文化》 18集

## 上海市嘉定区图书馆:

《嘉定文化故事》专题片 9集

## 上海市金山区图书馆:

《金山之恋：一镇一品》专题片 8集

## 上海市奉贤区图书馆:

《“奉贤故事”——上海奉贤民俗文化》专题片 10集

## 上海市崇明区图书馆:

《崇明民俗文化故事》专题片 10集

## 东方社区信息苑:

《上海说唱曲艺系列》专题片 5集

《崇明扁担戏》专题片 1集

## 复旦大学图书馆:

诗笺92件，诗文410首，涉及作者40位

## 上海韬奋纪念馆:

邹韬奋相关数据 邹韬奋的生平介绍及图片、生活书店介绍及图片、邹韬奋家世及相关人物数据68条

生活书店大事记数据 150条

馆藏图书书目数据 1892条

韬奋年谱数据 1067条

# 创新点

Innovation

## 功能创新:

**丰富历史数据:** 结合先进技术与丰富历史数据;

**双重交互界面展示:** 通过时间轴和bigmap地图双重交互界面展示上海的历史风貌;

**多模态输出:** 支持文本、图片、视频等多种形式的展示, 增强用户的沉浸式体验。

## 技术创新:

**多模态神经网络:** 整合视觉和文本数据, 增强模型的信息处理能力;

**自然语言处理(NLP):** 提高用户交互的自然性和准确性;

**大规模知识图谱:** 提供结构化和关联化的历史数据支持;

**RAG技术:** 结合信息检索与生成式模型, 提高内容生成的准确性和相关性。

## 项目价值:

**教育与学术研究:** 为研究者提供丰富的数据和直观的历史洞察。

**文化保存与传播:** 提高公众对文化的认知与兴趣。

**多模态学习体验:** 结合文本、图像和视频提供多维度的学习体验, 使学习过程更加生动有趣。

**增强用户参与:** 高度互动的平台为用户提供一种全新的文化探索方式。

# 总结与反思

Reflection

## 不跑题，紧扣数据：

是开放数据比赛，数据要尽可能的利用到。

## 创新合理，尊重史实：

不能一味追求创新、增强用户体验的趣味性。  
要尊重文化史实，就算AIGC也要给出内容的历史依据、引用。

## 关注热点，把握风口：

大模型+RAG是今年上半年的风口，把握住，立足新颖。



谢谢观看

THANK

YOU