自然语言处理

搜索



会员中心 🞁 消息



# F源模型应用落地-LangSmith试炼-入门初体验-数据集评估(三)

## 、前言

LangSmith是一个用于构建生产级 LLM 应用程序的平台,它提供了调试、测试、评估和监控基于任何 LLM 框架构建的链和智能代理的功能,并能与LangC 使用LangSmith帮助开发者深入了解模型在不同场景下的表现,让开发者能够更高效地进行模型相关的开发、调试和管理。

本篇将学习 Evaluate 功能,主要用于评估和衡量在多样化数据上的性能和完整性。它提供了集成的评估和追踪框架,允许用户检查回归问题、比较系统, 吴来源及性能问题。

## 、术语

# 1.LangChain

是一个全方位的、基于大语言模型这种预测能力的应用开发工具。LangChain的预构建链功能,就像乐高积木一样,无论你是新手还是经验丰富的开发者, 的部分快速构建项目。对于希望进行更深入工作的开发者,LangChain提供的模块化组件则允许你根据自己的需求定制和创建应用中的功能链条。

LangChain本质上就是对各种大模型提供的API的套壳,是为了方便我们使用这些 API,搭建起来的一些框架、模块和接口。

#### angChain的主要特性:

- 1.可以连接多种数据源,比如网页链接、本地PDF文件、向量数据库等
- 2.允许语言模型与其环境交互
- 3.封装了Model I/O (输入/输出)、Retrieval (检索器)、Memory (记忆)、Agents (决策和调度)等核心组件
- 4.可以使用链的方式组装这些组件,以便最好地完成特定用例。
- 5.围绕以上设计原则, LangChain解决了现在开发人工智能应用的一些切实痛点。

### 2.LangSmith

是一个用于构建生产级 LLM 应用程序的平台,它提供了调试、测试、评估和监控基于任何 LLM 框架构建的链和智能代理的功能,并能与 LangChain 无缝活:

- ,调试与测试:通过记录langchain构建的大模型应用的中间过程,开发者可以更好地调整提示词等中间过程,优化模型响应。
- ,评估应用效果:langsmith可以量化评估基于大模型的系统的效果,帮助开发者发现潜在问题并进行优化。
- · **监控应用性能**:实时监控应用程序的运行情况,及时发现异常和错误,确保其稳定性和可靠性。
- ,**数据管理与分析**:对大语言模型此次的运行的输入与输出进行存储和分析,以便开发者更好地理解模型行为和优化应用。
- · **团队协作**: 支持团队成员之间的协作, 方便共享和讨论提示模板等。
- · 可扩展性与维护性: 设计时考虑了应用程序的可扩展性和长期维护, 允许开发者构建可成长的系统。

## 3.LangChain和LangSmith的关系

LangSmith是LangChain的一个子产品,是一个大模型应用开发平台。它提供了从原型到生产的全流程工具和服务,帮助开发者构建、测试、评估和监控也 LLM 框架的应用程序。

LangSmith与LangChain 的关系可以概括为: LangChain是一个开源集成开发框架,而 LangSmith是基于LangChain 构建的一个用于大模型应用开发的平

### 、前提条件

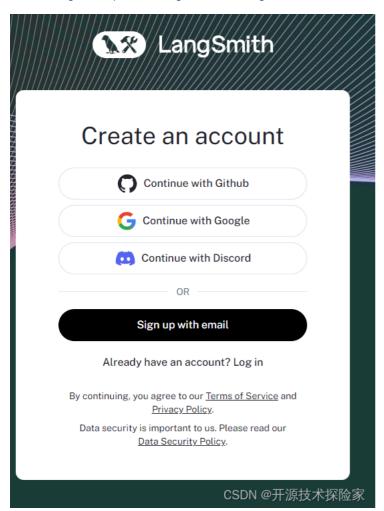
## 1.安装虚拟环境

- 1 conda create --name langsmith python=3.10
- 2 conda activate langsmith
- 3 | pip install -U langsmith -i https://pypi.tuna.tsinghua.edu.cn/simple
- 4 pip install openai -i https://pypi.tuna.tsinghua.edu.cn/simnle

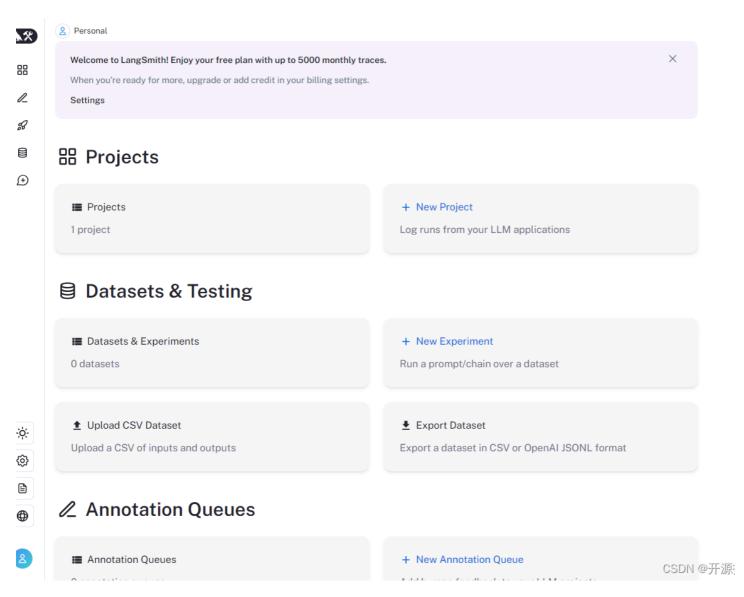


# 2.创建API key

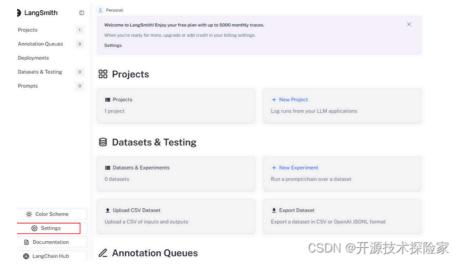
作入口: LangSmithhttps://smith.langchain.com/settings未登录的需要先进行登录:



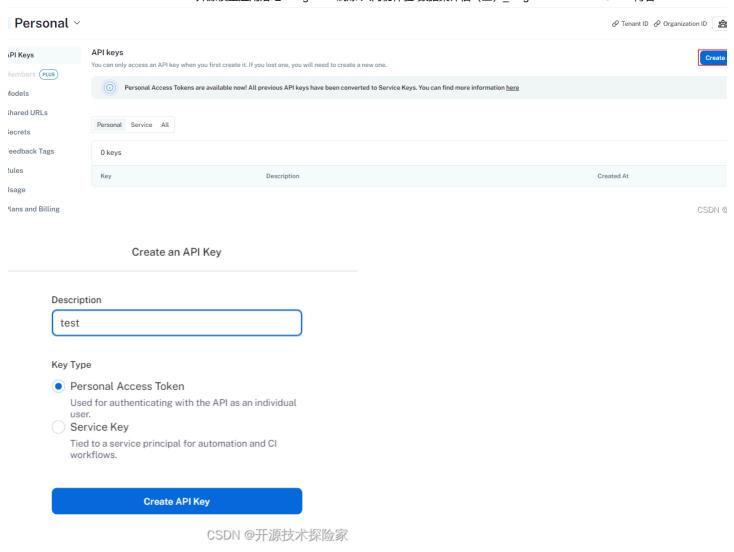
录成功:



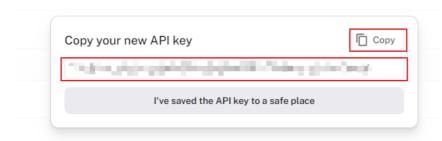
### 击Settings:



### 击Create API Key:



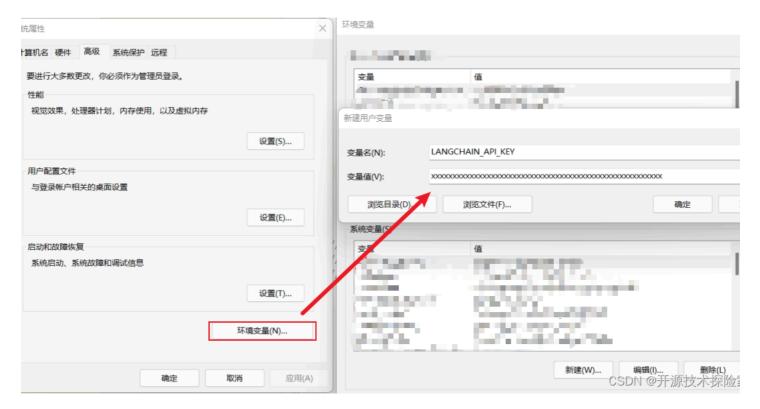
### 录API Key:



CSDN @开源技术探险家

# 3.设置环境变量

ndows:



ux:

需要替换3.2创建的API Key

## 4.生成样例数据

义使用GPT去生成,例如:输入"<u>请生成10条酒店的评论,需要包含好评和差评的标识,数据以JSON格式返回,具体参见示例:</u> <u>("comment":"这个酒店设</u>施 <u>下次都不来了","reviews":"差评"}</u>"

6/16

开源模型应用落地-LangSmith试炼-入门初体验-数据集评估(三)\_langsmith.schemas-CSDN博客

请生成10条酒店的评论,需要包含好评和差评的标识,数据以JSON格式返回,具体 参见示例: {"comment":"这个酒店设施太老旧,房间有异味,下次都不来 了","reviews":"差评"}

Assistant Poe

```
好的,以下是10条酒店评论,包含好评和差评的标识,JSON格式:
"comment": "这家酒店环境很好,交通便利,服务员非常友好,下次还会来!",
"reviews": "好评"
"comment": "房间虽小但很干净整洁,性价比很高,下次还会选择这里。",
"reviews": "好评"
"comment": "酒店的游泳池和健身房设施一流,住宿体验非常棒,下次有机会一定再来。",
"reviews": "好评"
},
"comment": "早餐种类很丰富,味道也不错,服务态度很好,对得起这个价格。",
"reviews": "好评"
},
"comment": "酒店位置非常优越,临近著名景点,出行非常方便,下次有机会一定会再入住。",
"reviews": "好评"
"comment": "房间设施有些陈旧,卫生情况一般,不太符合我的预期,下次不会再选择这里。",
"reviews": "差评"
},
{
```

与Assistant交谈或@一个机器人



```
1 [
2
   "comment": "这家酒店环境很好,交通便利,服务员非常友好,下次还会来!",
3
4
   "reviews": "好评"
5
  },
   "comment": "房间虽小但很干净整洁,性价比很高,下次还会选择这里。",
   "reviews": "好评"
9
LØ
  {
```

## 、技术实现

### 1.准备数据集

```
1 | # -*- coding = utf-8 -*-
2 import os
3
4
  from langsmith import Client
  from langsmith import schemas as ls_schemas
6
```



```
7 | os.environ['LANGCHAIN_TRACING_V2'] = 'true'
                     LØ
 if __name__ == '__main__':
```

#### 刊结果:

Comment Queries ky 0 10 2024/5/25 21:31:19
--

CSÓN @剂

periments Pairwise Experiments Examples						
Filters	Base Split All +					
	Input	Output	Created At	Modified At	Splits	
	这家酒店环境很好,交通便利,服务员非	好评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	房间虽小但很干净整洁,性价比很高,下	好评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	酒店的游泳池和健身房设施一流,住宿	好评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	早餐种类很丰富,味道也不错,服务态度	好评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	酒店位置非常优越,临近著名景点,出行	好评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	房间设施有些陈旧,卫生情况一般,不太	差评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	服务员的态度不太友好,办理入住和退	差评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	酒店停车场收费很高,且位置不太好找,	差评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	房间隔音效果很差,能听到走廊和邻居	差评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
	酒店餐厅的菜品种类和口味都一般,性	差评	2024/5/25 21:31:20	2024/5/25 21:31:20	base	
examples i	n total				〈 〉 形® NdSD	

# 2.定义目标任务

```
1 openai = wrappers.wrap_openai(Client())
2
3
      @traceable
      def label_text(text):
4
5
          messages = [
6
7
                  "role": "system",
8
                  "content": "请分析下面的用户评论,并确定它是积极还是消极。如果是积极,用"好评"回答,如果是消极,用"差评"回答。",
9
              },
LØ
              {"role": "user", "content": text},
L1
          ]
L2
          result = openai.chat.completions.create(
L3
              messages=messages, model="gpt-3.5-turbo", temperature=0
L4
L5
          return result.choices[0].message.content
```

## 3.配置评估器对输出进行评分,运行评估:



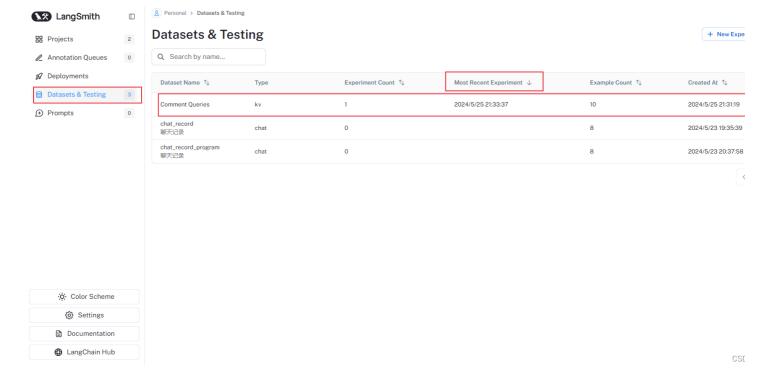
```
1 # 配置评估器对输出进行评分
2
   def correct_label(root_run: Run, example: Example) -> dict:
       score = root_run.outputs.get("output") == example.outputs.get("reviews")
3
4
       return {"score": int(score), "key": "correct_label"}
5
6
7
   dataset_name = "Comment Queries"
8
9
   # 运行评估并查看结果
LØ
   results = evaluate(
```

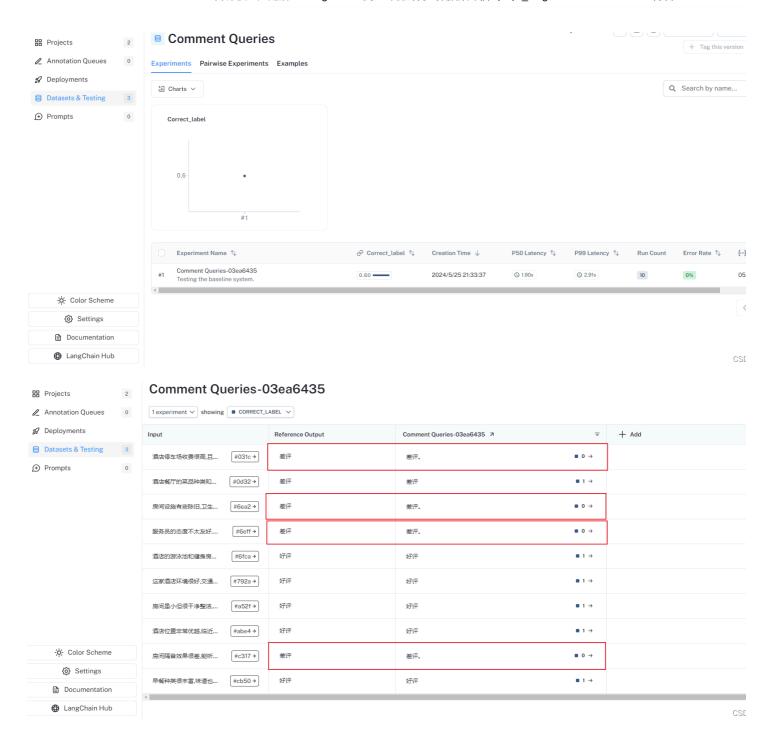
#### 刊结果:

### EA输出:



#### ngSmith 控制输出:





仑:

面运行的10个样本数据中,有4个模型的输出结果跟预期不一致(<del>期望输出"差评",实际输出"差评。"</del>),所以正确率为60%。

### 4.代码优化

上一步评估测试后,正确率只有60%,接下来,尝试优化代码,再次进行评估。

们修改System Prompt,改为:"请分析下面的用户评论,并确定它是积极还是消极。如果是积极,只需要回答"好评"两字,如果是消极,只需要回答"差评"

```
1
   @traceable
2
   def label_text(text):
3
      messages = [
4
         {
5
             "role": "system",
             "content": "请分析下面的用户评论,并确定它是积极还是消极。如果是积极,只需要回答"好评"两字,如果是消极,只需要回答"差评"两字。",
6
7
         },
8
          {"role": "user", "content": text},
9
      ]
LØ
      result = openai.chat.completions.create
                                              开源技术探险家 ( 关注 )
                                                                                                         ♣ 47 ♣ ★ (
L1
          messages=messages, model="gpt-3.5-t
```

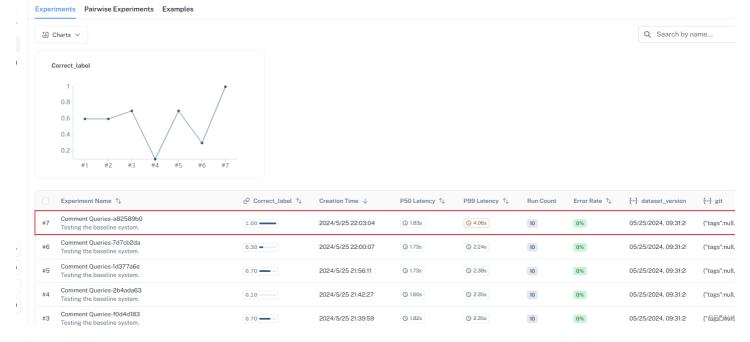
13  $\verb"return result.choices" [ \verb"o"] . \verb"message.content" \\$ 

### 刊结果:

#### EA输出:



#### ngSmith 控制输出:



### Comment Queries-a82589b0

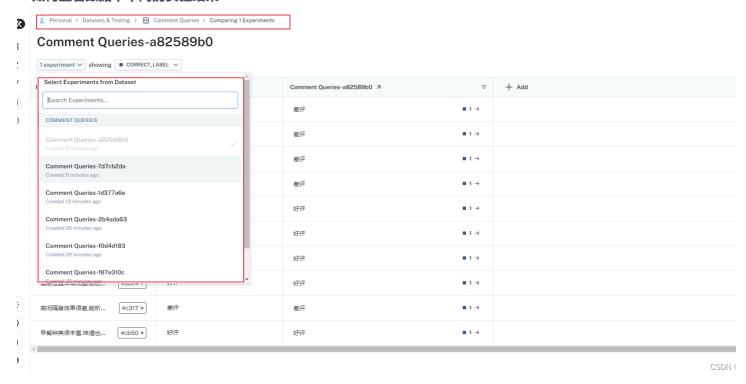
?	1 experiment v showing 0 CORRECT_LABEL v					
7	Input	Reference Output	Comment Queries-a82589b0 7	Ŧ	+ Add	
÷	酒店停车场收费很高,且 #031c →	差评	差评	■ 1 →		
	酒店餐厅的菜品种类和 #0d32 →	差评	差评	■ 1 →		
	房间设施有些陈旧,卫生 #6ea2 →	差评	差评	■ 1 →		
	服务员的态度不太友好, #6eff→	差评	差评	■ 1 →		
	酒店的游泳池和健身房 #6fca →	好评	好评	■ 1 →		
	这家酒店环境很好,交通 #792a →	好评	好评	■ 1 →		
	房间虽小但很干净整洁, #a52f →	好评	好评	■ 1 →		
	酒店位置非常优越,临近 #abe4 >	好评	好评	■ 1 →		
	房间隔音效果很差,能听 #c317 →	差评	差评	■ 1 →		
	早餐种类很丰富,味道也 #cb50 →	好评	好评	■ 1 →		
þ					) MD20	

## 仑:

面运行的10个样本数据中,所有输出结果跟预期一致,准确率100%。

# 附带说明

### 1. 如何查看数据下不同的实验结果



## 2. 数据集备选方案

了按照4.1.准备数据集以外,还可以选用已有的数据集,具体参见:

原模型应用落地-LangSmith试炼-入门初体验-数据集管理(二)-CSDN博客

### 3. 完整代码

```
1 | # -*- coding = utf-8 -*-
2
  import os
  from langsmith import wrappers, traceable
  from langsmith.schemas import Example, Run
5
  from langsmith.evaluation import evaluate
6
  from openai import Client
7
  8
9
  os.environ['LANGCHAIN TRACING V2'] = 'true'
10
```

## 文章知识点与官方知识档案匹配,可进一步学习相关知识

vthon入门技能树 首页 概览 433880 人正在系统学习中

原模型应用落地-LangSmith试炼-入门初体验-追踪代码 (一) 索如何使用LangSmith实现更高效地进行模型相关的开发、调试和管理 没有卑微的工作,只有卑微的心态,与其

#### 大模型全栈工程师课程笔记 - LLM tools

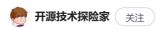
Michael是

xort repass@property@property@property# 正则表达式定义了要去除的标点符号self,) -> dict:# 自定义的BLEU SCORE评估器。

**}评论** 



ss7424 🙏 有个问题, messages=messages, model="gpt-3.5-turbo", temperature=0,这段代码不就是基于GPT来判断自己的LLM生成的数据集来进行扩展。





### ngChain 77 LangSmith 从入门到精通二 langsmith如何使用

nlangsmith.evaluationimportEvaluationResult,run\_evaluatorfromlangsmith.schemasimportExample,Run@run\_evaluatordefcheck\_not\_idk(run:Run,example:Example):"""llustra

### ngChain(0.0.339)官方文档二:LCEL

N和輸出模式:輸入和輸出模式(schemas)为每个LCEL链条提供了基于您链条结构推断的Pydantic和JSONSchema模式,也就是可以根据您链条的结构自动生成数据模式,确保数据信息

#### ngChain 与 LangSmith:构建与微调支持LLM的智能应用双重攻略

盲人工智能领域的不断发展,<mark>自然语言处理</mark> (NLP) 技术也在不断进步。近年来,革命性的大型<mark>语言模型</mark> (LLM) 已经成为了 NLP 技术的重要组成部分。作为一种可以理解并生成刻

#### ngSmith 入门

听的页面需要登录,然后点击头像,然后是Settings。第一行API keys,然后Create Api Key。设置环境变量(windows10)查看LangSmith输出轨迹。

#### 模型LangChain框架基础与使用示例

天的社区与生态:Langchain团队迭代速度非常快,能快速使用最新的语言模型特性,该团队也有 langsmith, auto-evaluator 等其它优秀项目,并且开源社区也有相当多的支持。LangC

#### D(Schema)教程 schema.xsd 如何编辑和添加新内容

可使用 XML Schemas? XML Schema is an XML-based alternative to DTD. XML Schema是以XML语言为基础的,它用于可替代DTD。 An XML schema describes the structure o

#### ngChain 18 LangSmith监控评估Agent并创建对应的数据库

了记录运行,<mark>LangSmith</mark>还允许您测试和<mark>评估</mark>LLM<mark>应用</mark>程序。在本节中,您将利用<mark>LangSmith</mark>创建基准<mark>数据集,</mark>并在代理上运行AI辅助<mark>评估</mark>器。

#### ngChain系列使用指南: LangChain快速入门

门仅涉及了提示、<mark>模型</mark>和输出解析器的基础知识 - 要深入了解这里提到的所有内容,请参阅文档的此部分。我们只是简单介绍了检索的基础知识 - 想要深入了解这里提到的所有内

#### ngChain教程 | LCEL原理详解教程 — | LangChain Expression Language...

N和輸出模式:輸入和輸出模式为每个LCEL链提供了从链的结构中推断出的Pydantic和JSONSchema模式。这可用于验证输入和输出,是LangServe不可或缺的一部分。 无缝Lang

#### ngChain Model I/O解读 lang chain model

ngChain Model I/O解读 一 总览 模型IO的目标是使开发者无须深入理解各个模型平台的API调用协议就可以方便的与各个大模型进行交互。 概念指南 对消息、提示、LLM与聊天

#### 原模型应用落地-语音转文本-whisper模型-AIGC应用探索(四)-测试音频

原模型应用落地-语音转文本-whisper模型-AIGC应用探索(四)-测试音频

#### 原模型应用落地-语音转文本-whisper模型-AIGC应用探索(二)-测试音频

原模型应用落地-语音转文本-whisper模型-AIGC应用探索 (二) -测试音频

## ngChain 36 深入理解LangChain 表达式语言优势— LangChain Expression...

Input and output schemas 输入和输出模式 输入和输出模式根据您的链的结构推断出每个LCEL链的Pydantic和JSONSchema模式。这可用于验证输入和输出,并且是LangServe

#### angchain】入门初探实战笔记(Chain, Retrieve, Memory, Agent) la...

agsmith can generate test cases based on the specificationsorrequirements provided. These test cases can cover various combinations of inputsandexpected outputs, ensuring

### 3HELL-1-sample数据集

3HELL-1-sample<mark>数据集</mark>是专为中文语音识别技术研发设计的一个<mark>开源数据集</mark>,它在AI和语音处理领域中有着广泛的<mark>应用</mark>。这个<mark>数据集</mark>由北京阿里云公司制作,旨在为学术界和工

#### ar-10数据集

ar-10数据集是计算机视觉领域广泛使用的开源数据集,主要由Alex Krizhevsky、llya Sutskever和Geoffrey Hinton等人创建。这个数据集的全名是"Canadian Institute For Advance

## ısorflow-dataset:IONDV应用程序用于形成和标记数据集以及基于数据集检查不同设置的张量流模型训练的结果

isorflow<mark>数据集应用程序-是基于的应用</mark>程序 无需进行标准功能编程即可累积数据,标准化和标记图像,创建,训练和比较<mark>模型</mark>的()。还可以使用Node.js代码开发以<mark>模型</mark>修改和

## 你本地化部署与使用一款免费的LLM应用工程化平台

m0 631

自已经就绪,现在需要让<mark>应用与Lan</mark>gfuse集成起来,以能够跟踪、<mark>评估</mark>与分析我们的<mark>应用。</mark>使用Langfuse的SDK/API,Langfuse能够与现有的任意<mark>应用</mark>做集成,但如果你的<mark>应用</mark>。

# igsmith功能总结

daada

介段下的langsmith功能总结

Ms之LangChain: LangChain(开发和构建)、LangSmith(生产化/跟踪)、LangServe(部署为API)的基础使用案例教程—... 近期请国内外头部出版社可尽快私信博主 vls之LangChain: LangChain(开发和构建)、LangSmith(生产化/跟踪)、LangServe(部署为API)的基础使用案例教程—利用LangChain开发和构建LLM的应用程序(构建LLM链/检算

#### 于深度学习的点云降噪 最新发布

F<mark>深度学习</mark>的点云降噪是一种利用<mark>深度学习模型</mark>处理三维点云数据,以去除噪声并恢复点云的原始形状和细节的方法。点云数据广泛<mark>应用于</mark>自动驾驶、机器人导航、3D扫描和虚

### ansformer-LSTM分类 | Matlab实现Transformer-LSTM多特征分类预测/故障诊断

关注并私信文章链接, 获取对应文章源码

nsformer-LSTM分类 | Matlab实现Transformer-LSTM多特征分类预测/故障诊断

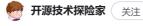
## 深度学习】图形模型基础(6):模型优化理论

MI

<mark>坚</mark>优化理论关注提升<mark>模型</mark>性能、准确性和效率。SGD(随机梯度下降)是其关键算法之一,通过随机选取样本计算梯度并更新<mark>模型</mark>参数,以提高训练速度和适用性。SGD尤其适

### 工智能及深度学习在病理组学中的应用概述 | 系列推文·24-07-11

七,小罗会带大家回顾计算机和编程的起源,解释从最初的电子计算器到现代计算机的发展历程,以及高级编程<mark>语言</mark>如何简化计算机指令的编写。其次,本期推文会深入探讨机。





### 可开发一个java开源框架-----Jvn入门篇

安一个Java<mark>开源框架需要掌握以下几个步骤: 1.明确框架的目标和使用场景。 在开发框架之前,需要明确框架的目标和使用场景。例如,框架是用来简化开发、提高代码可复用</mark>

### "相关推荐"对你有帮助么?

非常没帮助









关于我们 招贤纳士 商务合作 寻求报道 ☎ 400-660-0108 ☑ kefu@csdn.net 👨 在线客服 工作时间 8:30-22:00

公安备案号11010502030143 京ICP备19004658号 京网文〔2020〕1039-165号 经营性网站备案信息 北京互联网违法和不良信息举报中心 家长监护 网络110报警服务 中国互联网举报中心 Chrome商店下载 账号管理规范 版权与免责声明 版权申诉 出版物许可证 营业执照 ©1999-2024北京创新乐知网络技术有限公司





私信

关注

Q



## 文章

!博主文章

模型应用落地-工具使用篇-Ollama

) 0 19793

模型应用落地-qwen-7b-chat与vllm实 理加速的正确姿势 (一) ◎ 14711

模型应用落地-解锁大语言模型的无限

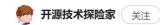
14011

模型应用落地-chatglm3-6b-streamlit-篇(八) ① 12732

模型应用落地-工具使用篇-SLB (二) 12430

### 专栏









#### 评论

模型应用落地-工具使用篇-Spring Al-... paotoutuo: 开源模型应用落地工具Sprin 系列, 这篇Function Call (八) 真不 ...

模型应用落地-FastAPI-助力模型交... 电子空间: 干货满满, 实用性强, 博主 作风格简洁明了,让人一目了然。3 ...

模型应用落地-音乐生成模型-suno/b... hk: 这是一篇非常实用的技术博客

模型应用落地-LangChain高阶-智能... 忘情: 这篇文章的结构非常清晰,让读者 轻松地跟随作者的思路。这种条理清...

模型应用落地-音乐生成模型-openai/... ıgit02: 这种富有创意的观点值得进一步 探讨和研究。

### 意向朋友推荐"博客详情页"吗?













不推荐 不推荐 一般般

推荐 强烈推荐

#### 文章

模型应用落地-Yi模型小试-闭源模型-yi-⇒LangChain (五)

模型应用落地-Yi模型小试-开源模型 vs 模型 (四)

模型应用落地-Yi模型小试-Yi-1.5-9Bt-16K集成vLLM-入门篇 (三)

4

7篇

7月	06月	05月	04月
2篇	22篇	28篇	25篇
3月	02月	01月	

22篇

12篇

### 前言

## 术语

- 2.1.LangChain
- 2.2.LangSmith
- 2.3.LangChain和LangSmith的关系

## 前提条件

- 3.1.安装虚拟环境
- 3.2.创建API key
- 3.3.设置环境变量
- 3.4.生成样例数据

## 技术实现

- 1.1.准备数据集
- 1.2.定义目标任务

