

Investigation mate

Wednesday, April 23, 2025 09:34

<https://github.paypal.com/Compliance-R/compmediasearchserv>

<https://github.paypal.com/jshi3/investigation-agent/>
<https://paypal.atlassian.net/wiki/spaces/CIPDS/pages/2231620414/Investigation+Mate>

<https://paypal.atlassian.net/wiki/spaces/CIPDS/pages/2238986819/Task+Breakdown>

<https://paypal.atlassian.net/wiki/spaces/GDGRTPlatform/pages/574409076/BRM-Adult+Typology+Indicator>
<https://paypal.atlassian.net/wiki/spaces/GDGRTPlatform/pages/574400521/Adult+typology+pilot+one+page>

Adult Typology FYR

.env

AUTHORITY= <https://login.microsoftonline.com/fb007914-6020-4374-977e-21bac5f3f4c8>

TENANT_ID=fb007914-6020-4374-977e-21bac5f3f4c8

GRANT_TYPE=client_credentials

CLIENT_ID=57379ce5-d853-4bfa-9f8e-53194ad39ace

CLIENT_SECRET=

SCOPE= <https://cognitiveservices.azure.com/.default>

AZURE_OPENAI_ENDPOINT= <https://genai-cus-tdz.paypalcorp.com/adverse-media-search>

AZURE_OPENAI_API_TYPE=azure_ad

AZURE_OPENAI_API_VERSION=2024-02-15-preview

AZURE_OPENAI_CHAT_COMPLETIONS_DEPLOYMENT_ID=gpt-4o_2024-05-13

AZURE_OPENAI_CHAT_COMPLETIONS_MODEL=gpt-4o

---Agent LLM Config---

AGENT_LLM_TYPE=azure_gpt_4o

#AGENT_LLM_TYPE=cosmos

#AGENT_LLM_MODEL=deepseek-r1-dis-3ad7d

#AGENT_LLM_TYPE=ollama

#AGENT_LLM_MODEL=qwen2.5-coder:32b

---Formatting LLM Config---

FORMATTING_LLM_TYPE=azure_gpt_4o

#FORMATTING_LLM_TYPE=cosmos

#FORMATTING_LLM_MODEL=deepseek-r1-dis-3ad7d

#

#FORMATTING_LLM_TYPE=ollama

#FORMATTING_LLM_MODEL=qwen2.5-coder:32b

---DATA_ANALYZER and DATA_EXTRACTOR LLM Config---

DATA_ANALYZER_LLM_TYPE=azure_gpt_4o

```
#DATA_ANALYZER_LLM_TYPE=cosmos
#DATA_ANALYZER_LLM_MODEL=deepseek-r1-dis-3ad7d
##
#DATA_ANALYZER_LLM_TYPE=ollama
#DATA_ANALYZER_LLM_MODEL=qwen2.5-coder:32b

# ---TRANSACTION NOTE SUMMARY LLM Config--- #
#TRANSACTION_NOTE_SUMMARY_LLM_TYPE=azure_gpt_4o

#TRANSACTION_NOTE_SUMMARY_LLM_TYPE=cosmos
#TRANSACTION_NOTE_SUMMARY_LLM_MODEL=deepseek-r1-dis-3ad7d
##
TRANSACTION_NOTE_SUMMARY_LLM_TYPE=ollama
TRANSACTION_NOTE_SUMMARY_LLM_MODEL=qwen2.5-coder:32b
```

CLIENT_SECRET=ir_8Q~IPjuGg25HvLZkjX2PdgQqYTpTrm2MTUbgE 可以填这个

是否行v要link

判断linked account 有phama

交易

Database url 同步异步 web inspection

F3b1a830954179eb25fff53fadc532dae7819cae

```
(1)获取DB API, curl --location 'https://aiplatform.dev51.cbf.dev.paypalinc.com/seldon/seldon/web-
inspection--3d643/api/v0.1/predictions' \
--header 'Content-Type: application/json' \
--data '{
  "data": {
    "text": [
      "advanced turmeric"
    ]
  },
  "authorization": {
    "token": "", "project_id": "dev52-test-apps-cipds"
  }
}
```

}', 从transaction note/ dispute / limitation/ email / paypal me 等文本当中提取出可能会violate aup的地方, 包括但不限于phama, 然后去db里根据keyword匹配, 得到匹配结果 keyword_match. (2)根据llm提供的phama related keyword样例, 去文本中抽取可能涉及phama的关键词, 作为 model_matches. (3) combined_matches = list(set(keyword_matches + model_matches))。整体的输入为dataAnalyzerTool 得到的list(某个sheet的某个column, 比如根据用户query 加载的是masked_casefile.xlsx 当中disputes sheet当中notes这一列)。输出为是否存在aup violation, 如果有则进一步输出aup_violation_categories及文本中violation的内容

1. Jingjing phama db
2. Jiongsheng API

~ 11:11:11

3. 中文注释
4. Unit test
5. 输入是list, 就是上一个tool输出的某个sheet的某个column
6. <https://github.paypal.com/iaod-platform/Web-Inspection-MCP-Server/blob/main/server.py> 涉及到Raptor参考这个

数据: query

代码: 已有

结果: 跑出结果即可

Url: API 的基础url 这里通过fstring构造, , 指向一个网检查服务的端点。Web_inspection_base_url 变量应该在代码的其他地方定义, 代表网页检查服务的基础地址。例如, 它可能是 <https://api.example.com>. 完整的url可能是

Method http请求方法, 这里是post, 意味着将使用post请求向api发送数据

query_params 一个空字典{} 表示在http请求中不包含任何查询参数

Headers 一个空字典{} 表示在http请求中不包含任何自定义头部

Timeout 请求超时时间, 设置为5秒

Body 请求体 包含要发送给api的数据, 这是一个字典, 包含了网页检查的配置信息:

Use_case 使用场景, 由web inspection use case 变量指定, 这可能表示检查的目的, 例如 compliance 或security

params: 一个参数列表, 包含实际的检查目标和设置

第一个参数的name是domain_urls, value 是 to_inspect_urls to_inspect_urls应该是一个包含待检查网页url的列表

第二个参数的name是crawl_setting, value是一个包含爬虫设置的字典:

Max_depth 爬虫的最大深度, 由crawler_max_depth 变量指定

Max_pages 爬虫的最大页面数, 由crawler_max_pages 变量指定

Enable_screenshot 是否启用截图, 由crawler_enable_screenshot 变量指定

enable_image_download 是否启用图片下载, 由crawler_enable_image_download 变量指定 (布尔值)

enable_javascript: 是否启用javascript执行, 由crawler_enable_javascript 变量指定 (布尔值)

source_service: 发送请求的服务, 其值在代码的其他地方定义

target_service: 目标服务, 即web_inspection_service, 其值也在别处定义

km_application_context: 与知识管理系统相关的应用上下文, 其值在别处定义

trigger_api_client: 创建好的RaptorApiClient 实例被赋值给变量trigger_api_client 之后, 可以使用这个对象来实际出发网页检查。例如:

Response= trigger_api_client.predict()

整体思路:

这段代码的目的是创建一个API客户端, 用于向一个网页检查服务发送请求。请求体包含了要检查的url列表以及爬虫的配置参数(如最大深度、页面数、是否启用截图等) RaptorApiClient 类封装了与API交互的细节, 使得调用者只需要提供必要的参数, 而无需关心底层的http请求构建和发送, 创建好客户端后, 可以调用其方法来实际触发检查

这段代码是网页检查流程中的一个关键步骤, 负责构建和配置API请求。它将检查目标url和爬

虫设置打包成一个请求，准备发送给后端的网页检查服务。

RaptorCallClient 类的使用提高了代码的可读性和可维护性，将API交互的复杂性隐藏起来

几个API调用组合来组合去

各种语言分布情况，就是在那个group by 语言，只看phama 和phama db

先确认下是否能申请

把这里的数据/Users/beibwang/Downloads/chatbot/investigation-agent/eval/predefined.yaml

即所有predefined questions

用最新aup violation代码

<https://github.paypal.com/jshi3/investigation-agent/pull/8/files>

跑一遍

尤其是

```
- vars:
  title: Email Address and PayPal.me Pharma Violation
  question: |-
    Highlight any pharmaceutical-related terminology or verbiage in the email address or
    PayPal.me slug of seed account.
- vars:
  title: Seed Account Limitation
  question: |-
    Check if there are any limitation on the seed account for each limitation type, especially
    Pharma/Drug related limitation.
- vars:
  title: Transaction Notes Pharma Violation
  question: |-
    List review accounts' last 3 years transaction details where violations related to
    pharmaceuticals can be identified via transaction notes, and provide drug product details from
    Pharma Info Database.
- vars:
  title: Disputes Comments Pharma Violation
  question: |-
    List review account's disputes information that indicate the sale of drugs/pharma from last
    3 years disputes data. Direct to each dispute case in SCM system
- vars:
  title: Transaction Notes Business Model
  question: |-
    Do analysis on Transaction Notes, whether can tell business model or no meaningful notes or
    mostly empty?
```

FPTI SDK 升级版本 raptor 玉环 service 端FPTI 升级版本号

中间有个步骤 要每个service加一个application context 需要application owner

Json type

PHARMA_DB_API_URL=https://aiplatform.dev51.cbf.dev.paypalinc.com/seldon/seldon/pharma-db-keywo-
ad1d9/api/v0.1/predictions
PHARMA_INFO_API_URL=https://te-iaodplatform.qa.paypal.com:21777/galaxi/api/v1/brm/pharma-info/

主动写ppt跟进完成情况
主动约会议写代码

名牌 去美国 电脑寄回来 不影响online 找

每周来个一两天
6.20号
Fully remote
在家 去美国
不要说去美国
就在家 不用说太死 就remote比较多 不要讲成fully remote
当然是到30号比较合适
跟team说一下

我肯定尽快去美国 3个月之后找工作

也可以提前结束
提前一周transfer交接工作
Remote 不说也可以, 电脑带出去不太方便, 不要打草惊蛇, 把事情做完就行了, 不要跟老板说就行了
明天周五答复
每天开会
在上海
跟铝材说一下 铝材同意 先推给tommy 海华; 也可以回去问问tommy跟海华, 然后定一下。我就去美国, 然后
fully remote, 6.30日last day
把事情做漂亮