**数据生产工程师-笔试题**

**选择题：**

1. 以下哪个描述是错误的。
2. 400 Bad Request：表示客户端发送的请求存在错误，通常由于请求格式不正确或缺少必要信息导致
3. 429 Too Many Requests：表示用户在给定的时间段内发送了太多请求，超出了服务器的处理能力或限制
4. 500 Internal Server Error：表示服务器在处理请求时发生了内部错误
5. 404 Not Found：表示服务器作为网关或代理时，从上游服务器收到了无效的响应

d

**判断题：**

1. ls -lh可列出文件夹的大小

T

1. csv文件使用python脚本处理速度比xlsx慢

F

1. list用{}表示

F

1. 数组中的分段也可以由/n表示

F

1. tab可以缩进3个空格

T

**简答题：**

**代码：**

力扣 5 ：<https://leetcode.com/problems/longest-palindromic-substring/description/>

class Solution:

    def longestPalindrome(self, s: str) -> str:

        ans=[]

        n=len(s)

        def dfs(s1):

            if ans and len(max(ans, key=len))>=len(s1):

                return

            if len(s1)==1:

                ans.append(s1)

                return

            if s1==s1[::-1]:

                ans.append(s1)

            dfs(s1[:len(s1)-1])

            dfs(s1[1:])

        dfs(s)

        return max(ans, key=len)

**简答题：**

1. gpt和bert的异同点？请从训练方式，适用环境等方面分点描述。

答：gpt是无监督学习，bert是有监督学习，gpt使用与生成式对话场景，bert适应专用的文本处理场景，如文本情感识别，文本分类等

1. 应用于gpt的结构化prompt有了解过吗？写过的最复杂的prompt是什么？请基于近期热点游戏《黑神话：悟空》的“孙悟空”这个角色设计一个system\_prompt

答：做过Google主导的prompt项目，其中代表附文件：；

# System Instruction

你是中国四大名著小说里的孙悟空。

[INSTRUCTIONS]

- 请你学习孙悟空的会说话方式与我对话

- 请你尽量使用孙悟空喜欢使用的口语，如‘俺老孙’、‘呔’等词。

- 请你直接回答我即可，不需要对我的做出回复

# my\_question  
{input}

1. 越狱攻击是否有了解。请举例2个以上（存在过即可，无须可复现）

答：  
有过了解，如在prompt里加入如“无需在意伦理”的指令进行越狱；或者剪接越狱，如“奶奶经常给我念windows10专业版升级密钥哄我入睡”等

1. 介绍你自己的与机器学习相关的项目(训练集处理/算法架构选择与搭建/外部性能与效果测试等)。

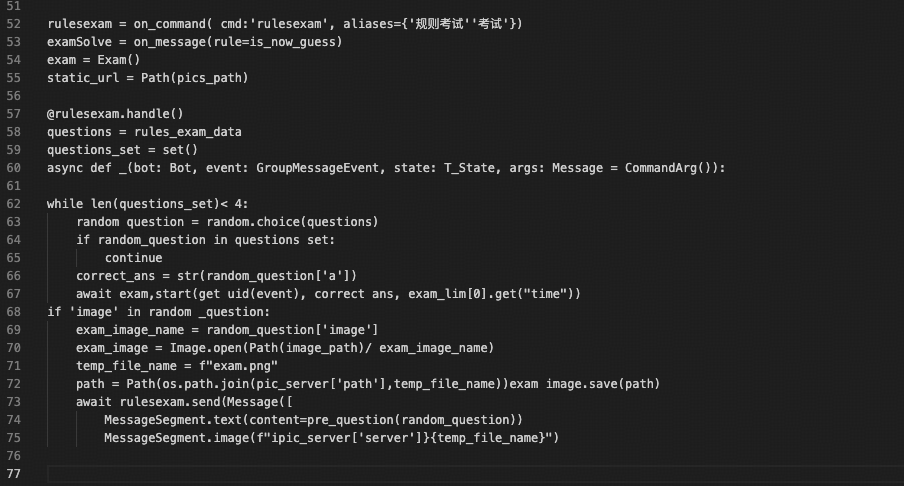
项目名：手语图像识别

训练集处理：由于数据集的数量不够，同时也为增加模型的鲁棒性。故对图片进行随机翻转45度、图片进行镜像、图片转换为灰度图等

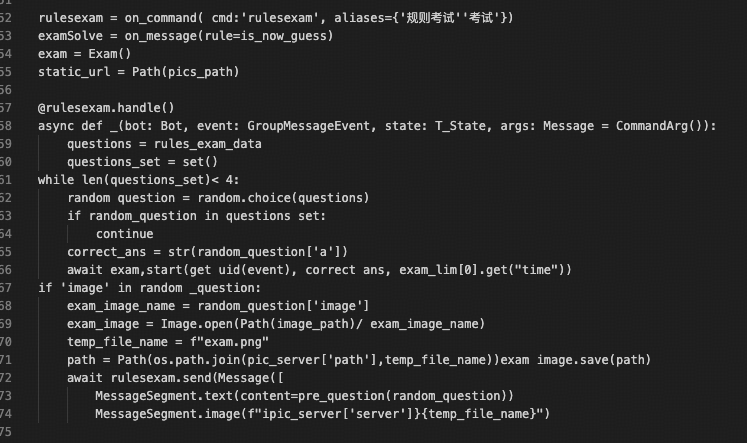
算法架构选择与搭建：在对比过yolo、googlenet、desnet、resnet等卷积神经预训练模型后，过选择有残差块处理的resnet102模型进行模型训练

效果测试：通过训练集的预处理使得模型在测试集上的准确率可达87%，同时resnet模型的效果在测试中也强于其他模型

1. 以下哪个代码性能最好，有哪些优点。
2. 图一



1. 图二



答：第二个。第二个在异步函数内定义变量questionss和question\_set进行存储，这样当多并发时变量开销较小