**数据处理代码：**  
①**transgpt内容提取.py**：从TransGPT官网下载下来的问答型数据json文件提取出表头为introduction和output的表格文件。将file\_path替换成合适的json文件就可以直接使用。

②**关键词提取.py**：提取transgpt问答数据里的“instruction”列或“output”列里含有关键词“轨道交通”的数据。

③**GPT-3.5-数据集产生.py**：使用GPT-3.5模型获取地铁领域的问答数据，设置输入模型的prompt为：“请你生成一条地铁领域的问答数据。输出格式为：问题：\*\*\*。回答：\*\*\*。”设置要生成的数据量total\_data = 2000。使用方法：

设置OpenAI的api之后就可以直接使用。

调用模型生成回答代码：

①**chatglm2-6b.py**：使用百度千帆大模型，设置好key，调用模型“ChatGLM2-6B-32K”，使用的prompt为“输出文本不要超过80个字。” + “问题”，这里的问题是测试文件里的问题列的问题数据。

②**ERNIE-3.5-8K.py**：使用百度千帆大模型，设置好key，调用模型：ERNIE-Bot-turbo，使用的prompt为“输出文本不要超过80个字。” + “问题”，这里的问题是测试文件里的问题列的问题数据。

③**llama.py**：使用百度千帆大模型，设置好key，调用模型：Qianfan-Chinese-Llama-2-7B，使用的prompt为“输出文本不要超过80个字。” + “问题”，这里的问题是测试文件里的问题列的问题数据。

④**通义千问.py：**调用模型：qwen-turbo，使用的prompt为“输出文本不要超过80个字。” + “问题”，这里的问题是测试文件里的问题列的问题数据。