MLE Take‑Home Assignment: Document Redlining with LLMs

*This assessment mirrors real‑world tasks: rule‑based detection, LLM integration, and performance evaluation.*

**Time:** *Up to 4 hours total (please limit yourself!)* \*\*\*\*

**API Credit:** *$50 LLM usage limit (API key to be shared over email)*

Good luck!

**About the Assignment:**

* Please **do not spend more than 4 hours** on this project.
* You will be provided with an **LLM API key** that has a **$50 usage limit** (shared via email)
* Submit your **code** and **output files** as part of your completed project. Submit **greater than 1hr before the live session you book below!**

**CRITICAL: Book a live interview to discuss this work within 4 days of receiving this document… see link**[**here**](https://calendly.com/nick-fleisher/greyball-interview-discussion)

**API Setup:**

1. Use the provided API key to access the LLM service.
2. Install any required Python packages.
3. Your code should run in a **standard Python 3.8+ environment**.

**Problem Statement**

This assignment is divided into two main tasks:

**1. Redlining Script**

Write a Python script that takes a document as input and outputs a list of issues (i.e., legal “redlines” or markups) based on a provided rule-based playbook.

* **Inputs:**
  + bad\_document.txt: A sample document that contains issues.
  + playbook.json: A JSON file describing the types of issues to detect and their definitions.
* **Outputs:**
  + A JSON file listing all identified issues in the following format:

[

{

"text\_snippet": "EXACT QUOTE",

"playbook\_clause\_reference": "EXACT CLAUSE NAME",

"suggested\_fix": "EXACT REPLACEMENT SUGGESTION"

},

...

]

* **Approach:**
  + Use the LLM API to process the document and detect issues based on the playbook.
  + You may use any prompt engineering techniques that help maximize accuracy.
  + Your script should be modular and clearly organized.

**2. Evaluation Script**

Write a second script to evaluate the redlines generated by your system.

* **Inputs:**
  + Your model's output (redline\_output.json)
  + The provided reference output (expected\_output.json)
* **Output:**
  + A summary of how well your system performed at identifying issues.
* **Evaluation Criteria:**
  + You are free to choose **any evaluation approach or library** you believe is appropriate.
  + The goal is to measure and summarize how closely your output aligns with the expected redlines.
  + Your script must run in a standard Python environment.

**Files Provided**

| **File Name** | **Description** | **Shared Link** |
| --- | --- | --- |
| bad\_document.txt | Document containing problematic language | <https://drive.google.com/drive/folders/1Z6uo0897LjJtT0b4_5kjLSz0fhQjd20H?usp=sharing> |
| playbook.json | Redlining rules/criteria | <https://drive.google.com/drive/folders/1Z6uo0897LjJtT0b4_5kjLSz0fhQjd20H?usp=sharing> |
| expected\_output.json | Reference list of correct redlines | <https://drive.google.com/drive/folders/1Z6uo0897LjJtT0b4_5kjLSz0fhQjd20H?usp=sharing> |
| llm\_api\_key.txt | API key for the LLM with $50 credit | To be shared via email directly |

**Deliverables**

| **Deliverable** | **File Name** |
| --- | --- |
| Redlining script | redline.py |
| Redlining output (JSON) | redline\_output.json |
| Evaluation script | evaluate.py |
| Evaluation summary/report | evaluation.txt |

Please make sure all scripts are self-contained and can be executed on a typical development machine.

**Requirements**

* Language: **Python 3.8+**
* You may use **any libraries** you prefer
* Your code must run in a **standard Python environment** without platform-specific dependencies.
* Stay within the **$50 API usage limit**.

**❓ FAQs**

**Q1. Can I use AI tools?**

A: Yes—feel free to iterate with AI, but we’ll discuss your reasoning live.

**Q2. What if I exceed the API credit?**

A: Aim to optimize usage; kindly note over‑consumption may affect cost fairness.

**Q3. How should I submit?**

A: Provide a GitHub repo link or zip with all scripts and outputs.

**Do not share any proprietary or confidential content from your existing or previous employers.**