Assignment Brief: LLM-powered Term Sheet Reconciliation

Context

As an Applied AI Developer in our Capital Markets division, your challenge is to create an automated reconciliation workflow. The goal is to compare key data points extracted from unstructured trade term sheets (in PDF/Word formats) with structured booking system records (CSV/JSON/XML). You will leverage a Large Language Model (LLM) that offers a public, free, or developer API to automate the parsing, extraction, and comparison. The final solution must be submitted as a public GitHub repository that is reproducible and well-documented.

Assignment Tasks

1. **Automated Document Parsing & LLM Extraction**
    - Parse the provided term sheets (PDF, Word) using Python.
    - Connect to an LLM via a free-tier or developer API (such as Perplexity, OpenAI with free credits/trial keys, Hugging Face, Cohere, etc.).
    - Automate the process so parsed document content is submitted to the LLM programmatically; extract all required fields (e.g., ISIN, coupon, maturity, issuer, currency, notional, settlement date).
    - No manual copy-pasting or web interface usage—all interactions must be performed by Python code/scripts via API.
2. **Automated Booking Extract Ingestion**
    - Parse the booking system extract (CSV/JSON/XML) using code.
    - Where necessary, upload/submit booking data for comparison via LLM, fully automated by script.
3. **Automated Reconciliation**
    - Use LLM output and your own logic to compare extracted term sheet details and booking records.
    - Generate a programmatic reconciliation table (CSV/Markdown/HTML), clearly showing matches/mismatches for each field.
4. **Documentation**
    - Include a concise report (max 2 pages) explaining your workflow, LLM/API integration, field extraction design, challenges, assumptions, and suggestions for additional automation or DevOps deployment in production.
5. **Repository and Submission**
    - Publish your complete solution as a public GitHub repository.
    - The repository must include:
        - All Python source code/scripts (well-structured).
        - Sample term sheets (PDF/Word), booking system extract files.

- A clear README.md with:
    - Step-by-step instructions to set up the environment (Python version, dependencies, API keys, etc.).
    - How to run each script end-to-end for a full workflow demonstration.
    - Configuration (e.g., API key setup instructions, required environment variables).
    - Description of expected input/output files/formats.
- The required documentation file.

6. **Testing & Reproducibility**
    - Your code should run end-to-end on the supplied mock data with just the setup steps described in your README.md.
    - We will download your repository and execute the workflow for review.

Constraints

- You must use an LLM that offers a free or public developer API. All LLM/API interactions must be fully automated—no manual steps.
- All deliverables must be in your public GitHub repository, ready to clone and run.
- You have 24 hours from receiving this assignment to complete and submit the work.