**Project Title:** Adverse News Screening for Financial Crime Surveillance

**Objective:** Develop a system to automatically screen publicly available news articles and identify entities (individuals or corporations) potentially involved in financial crime, scandals, or sanctions. This information will be used to enhance the bank's internal financial crime detection and prevention programs. The system should be able to identify and extract relevant entities from news articles, classify the nature of adverse news, and provide a confidence score for the relevance of the news to financial crime.

There are two types of adverse news search:

1. Broad-scale no target search within a time frame, e.g., past 1 month. This provide bank alerts to probe into its own customer base should some suspicious or convicted financial crime has been reported on bank’s customer before internal surveillance picks up the signal.
2. Targeted search where internal surveillance program triggered alerts on bank’s customer, and review team want to gather extra evidence from external media as supporting information to help investigation.

You may choose either one of the two types, or attempt both.

**Data Sources:**

* Candidates should identify and use publicly available news sources, such as:
  + Reputable financial news websites (e.g., Bloomberg, Reuters, Financial Times, Wall Street Journal).
  + News aggregators (e.g., Google News).
  + Sanctions lists and databases (e.g., OFAC, UN Sanctions).
  + Regulatory body websites (e.g., MAS, FCA, SEC).
  + Any other relevant publicly accessible sources.
* Candidates are encouraged to explore different data sources and justify their choices in their report. However, as many of such source are paid service, it is also fine to simply use google search API which provide sufficient free access limit.

**Task:**

1. **Data Collection and Preprocessing:**
2. **Entity Recognition and Extraction:**
   * Implement Named Entity Recognition (NER) to identify individuals and organizations mentioned in the news articles.
   * Develop a method to extract and disambiguate entities, ensuring that references to the same entity are grouped together (e.g., using techniques like coreference resolution or entity linking).
3. **Adverse News Classification:**
   * Train a machine learning model or leverage LLMs to classify news articles as containing adverse news related to financial crime, scandals, or sanctions.
   * Categories of adverse news can be pre-defined, e.g., Money Laundering, Terrorist Financing, Sanctions Violations, Fraud, Tax Evasion, Bribery and Corruption, Insider Trading, Ponzi and Pyramid Schemes, Trade-Based Money Laundering, etc., or it can be data-driven in an unsupervised learning way.
   * Explore different classification models, e.g., sentiment analysis, topic modelling, or more advanced deep learning approaches including directly leveraging LLM or gen AI services.
4. **Relevance Scoring:**
   * Propose a method to assign a confidence score to each identified adverse news item, indicating the likelihood that it is relevant to financial crime. Implementation is optional.
5. **Output and Visualization:**
   * Design a system to store and present the identified adverse news, including the extracted entities, the classification category, the relevance score, and the source of the news article. Implementation is optional.
   * Create visualizations to summarize the findings, such as trends in adverse news over time or the most frequently mentioned entities.

**Deliverables:**

1. **Python Code:** Well-documented and organized Python code for all stages of the project, upload to GitHub. Please include python version and library requirement.
2. **Presentation:** presentation slides (maximum 15 pages) covering the following aspects:
   * **Data Sources:** Description of the chosen data sources and justification for their selection.
   * **Methodology:** Detailed explanation of the approach used for entity recognition, adverse news classification, and relevance scoring. Include the specific models and techniques used, and the rationale behind their choice.
   * **Evaluation:** Discussion of the evaluation metrics used to assess the performance of the system (e.g., precision, recall, F1-score for classification). Provide quantitative results and analysis.
   * **Challenges:** Description of any challenges encountered during the project and how they were addressed.
   * **Business Value:** Discussion of the potential business value of the system for financial crime surveillance and detection. How can this system be integrated into a bank's existing compliance framework?
   * **Future Enhancements:** Suggestions for future improvements or extensions of the system.
3. **Presentation:** we may request a short presentation (15-20 minutes) by the shortlisted candidates to summarize the project, key findings, and business implications, followed by Q&A, in a 2nd round interview.

**Timeline:** please submit your completed project by the deadline specified in email.

**Tip**: Please use modern AI (ChatGPT, Gemini, Claude …) as much as it helps in developing the solution (cursor, cline) , as well as use it as part of the solution where applicable.

**Submission**: Please email your solution by replying the hiring manager’s email directly before the deadline in the email. (code please drop to GitHub as corporate email firewall sometimes block source code)