**Discussion 1.1 Compliance - First Documentation for an Open Source Software ML Pipeline**

1. **Sufficiency & Placement:** For a pipeline using a CC0 dataset (Kaggle) and an Apache-2.0 model (FinBERT), are README.md, LICENSE, and CITATION.cff alone sufficient? If not, where else should attribution and license notices appear (e.g., notebook headers, LICENSES/ folder, model weight metadata, data processing scripts) and why?
2. **Derivatives & Redistribution:** If you fine-tune FinBERT and/or produce transformed datasets from the CC0 source, what attribution and notice obligations persist for each derivative artifact (weights, exported models, processed data)? How should you document “state of modification” and provenance so downstream users remain compliant?
3. **License Strategy & Compatibility:** Choose a project license for your code (e.g., MIT vs. Apache-2.0 vs. GPL-3.0) given dependencies on an Apache-2.0 model and a CC0 dataset. Justify the choice with respect to compatibility, patent grants, NOTICE requirements, and future distribution of trained weights or packaged artifacts.