### **✅ 1. Instruction:**

**Review the core requirements for the overall theory, design, and selection of the modeling framework.**

**Relevant Context:**

*"LightGBM was selected due to its balance of interpretability and performance on tabular TF-IDF data. It provides feature importance outputs and rapid retraining."*

**Relevancy Score:** 5  
 ✅ Directly explains the rationale behind the model selection in alignment with theoretical and design requirements.

### **✅ 2. Instruction:**

**Assess whether the model assumptions are clearly stated and validated.**

**Relevant Context:**

*"Assumes independence between TF-IDF feature dimensions. Assumes the label annotations are accurate representations of ground truth."*

**Relevancy Score:** 4  
 ✅ Assumptions are stated clearly, but there's no explicit discussion of validation/testing of assumptions.

### **✅ 3. Instruction:**

**Review and summarize the information on data sources, including reasonableness of accessed or excluded data.**

**Relevant Context:**

*"The dataset includes 2,000 anonymized bankers’ notes extracted from the internal CRM platform... Each note was manually labeled... by trained compliance officers."*

**Relevancy Score:** 5  
 ✅ Data source origin and labeling process are clearly described, showing the data is internal, labeled by experts, and spans a relevant timeframe.

### **✅ 4. Instruction:**

**Review the data quality analysis performed in the MDD.**

**Relevant Context:**

*"No missing records were observed... Data was verified by compliance leads... All text records were processed using consistent UTF-8 encoding..."*

**Relevancy Score:** 5  
 ✅ Comprehensive coverage of completeness, accuracy, consistency, and preprocessing—aligns with standard data quality criteria.

### **❌ 5. Instruction:**

**Check for consistency between change logs and Model Metadata System regarding upstream models.**

**Relevant Context:** ❌ No mention of model change logs, **Model Metadata System**s, or upstream model dependencies.

**Relevancy Score:** 1  
 🚫 Not applicable in this MDD.

### **✅ 6. Instruction:**

**Assess whether the model outputs are consistent with the model objectives and requirements.**

**Relevant Context:**

*"The model is developed to classify bankers' notes as EFA or Non-EFA. It uses LightGBM as a binary classifier... trained on domain expert-labeled notes."*

**Relevancy Score:** 5  
 ✅ The output (binary label) clearly aligns with the stated business objective of identifying EFA language.

### **✅ 7. Instruction:**

**Evaluate the quality and comprehensiveness of the developer’s outcome testing.**

**Relevant Context:**

*"Performance metrics validated using 5-fold cross-validation. LightGBM outperformed Logistic Regression and Naive Bayes baselines by 6–8% F1 score."*

**Relevancy Score:** 4  
 ✅ Model evaluation is reported, but no deep dive into test coverage or specific business-oriented test cases.

### **❌ 8. Instruction:**

**Assess the adequacy of model implementation specifications and test plan.**

**Relevant Context:** ❌ No mention of implementation specifications, system test plans, or production deployment details.

**Relevancy Score:** 1  
 🚫 Missing implementation context.

### **❌ 9. Instruction:**

**Provide an assessment of model-level control activities documented in the MDD.**

**Relevant Context:** ❌ No explicit mention of model-level controls, governance, or risk mitigation mechanisms outside of preprocessing checks.

**Relevancy Score:** 2  
 🔍 Some implied data checks, but limited detail on broader control activities.

### **✅ 10. Instruction:**

**Assess the model monitoring plans (frequency, KPIs, thresholds, etc.).**

**Relevant Context:**

*"Sensitivity analysis showed stable predictions with +/-10% TF-IDF variance."* (But no ongoing KPI or threshold-based monitoring is described.)

**Relevancy Score:** 3  
 ⚠️ Some stability testing is included, but lacks full detail on formal performance monitoring plans, thresholds, or schedule.