**Automating iXBRL Tag Validation Using Generative AI**

**1. Problem Overview**

Currently, iXBRL tag validation is a heavily manual process. Reviewers use an Excel review log that lists every applied tag and must verify, line by line, whether each tag:

* Accurately represents the concept or financial fact it claims to
* Is correctly extracted from the financial statements
* Reflects the appropriate unit of measure and rounding
* Is applied to the right time period
* Contains proper dimensional context (e.g., differentiating between multiple directors)
* Is supported by narrative justification or commentary if the tag relates to textual content

This process is not only time-consuming and resource-intensive, but also error-prone, especially when dealing with lengthy financial documents and hundreds of tags. Mistakes in tagging or omissions can lead to regulatory non-compliance, rejected filings, and reputational damage.

**2. Proposed Solution Using GenAI**

To automate the iXBRL validation process, a hybrid system combining rules-based logic and large language models (LLMs) can be used for validation.

**Components of the System:**

1. **Input Files**
   * iXBRL Review Log (Excel): Contains all tags, values, units, periods, dimensions, and comments.
   * Financial Statements (PDF): Source document used for validation.
2. **PDF Section Parser**
   * Extract text from PDF.
   * Organize content into logical sections .
3. **Row-by-Row Tag Validation Engine**
   * For each tag in the Excel log:
     + **Numeric Tags**: Use regex and rule-based checks to validate presence, value, unit, and rounding.
     + **Narrative Tags**: Use an LLM to semantically match tagged content to source paragraphs in the PDF.
     + **Dimension Checks**: Ensure contextual accuracy (e.g., Director roles tagged correctly).
     + **Comment Validation**: Verify that explanations in the "Comments" column support tagging decisions.
4. **LLM-Powered Matching (Where Needed)**
   * Perform semantic similarity checks for narrative and contextual tags
   * Justify tagging decisions using LLM explanations
   * Suggest missed tags based on context
5. **Reporting Engine**
   * Generate a structured report listing:
     + Verified tags
     + Missing or inconsistent tags
     + Potential over-tagging
     + Unexplained omissions

**3. Key Fields in the iXBRL Excel Log**

Each field in the review log has a role in ensuring tag compliance:

* **Tag / Concept**: What is being tagged
* **Value**: The reported number or text
* **Unit**: Currency or measurement (e.g., EUR, %)
* **Accuracy**: Level of rounding applied (e.g., 1, 1000)
* **Period**: Time context for the tag
* **Dimension**: Adds specificity (e.g., which director)
* **Comments**: Explains tagging rationale or omissions

**Per-Tag Validation Checklist (Excel Row-Level Checks)**

Each tag (row in the Excel log) must undergo the following validations:

* **Tag Name (Concept)**: Confirm the tag correctly represents the reported financial or narrative item.
* **Value Accuracy**: Ensure the value exactly matches the corresponding figure or text in the PDF.
* **Unit**: Verify that the currency or measurement unit is consistent with the source document.
* **Accuracy (Rounding)**: Check that rounding (e.g., 1000s) is consistent with the PDF.
* **Period / Date Context**: Validate that the reporting period or instant date aligns with the document.
* **Dimension (Contextual Tagging)**: Ensure role-based tags (e.g., directors) use consistent and correct dimensions.
* **Narrative Match**: For textual tags, semantically match the description with its PDF source context.
* **Comments**: Review explanations for omissions, ambiguous tags, or overrides.