**Thesis Proposal: Recategorizing Incident Reports using ML or LLM: A Data Driven Approach**

**Introduction:**

Incident reporting is a critical process in various utility industries, including healthcare, aviation, and IT. Process safety incidents tend to happen infrequently but can often have catastrophic results when things go wrong[1]. Manual classification of incidents often leads to human errors, resulting in misclassification and inefficiencies. This thesis proposes the use of ML/LLM model to analyse and correct the classification of incident reports, leveraging historical data to improve accuracy and reliability.

**Literature Review**:

A review of existing literature will be conducted to understand the current state of incident reporting and classification. This will include studies on manual reporting errors, the application of ML/LLM in classification tasks, and the effectiveness of various algorithms in similar contexts.

**Problem Statement**:

Manual incident reporting is prone to errors due to **subjective judgement, fatigue, and lack of standardised criteria**. These errors can lead to **incorrect incident categorization**, affecting the **quality** of data and **subsequent decision-making processes**. The aim of this research is to develop a ML/LLM model that can **accurately classify incidents** and **correct** existing errors, **reducing human error** and **enhancing** the overall incident management process.

**Objectives**:

1. **Analyse the current manual incident reporting process** to identify common errors and inefficiencies.
2. **Develop a ML/LLM model** capable of accurately classifying incidents based on historical data.
3. **Evaluate the performance** of the machine learning model against manual classification.
4. **Implement the model** in a real-world setting to assess its impact on incident reporting accuracy and efficiency.

**Methodology**:

**Machine Learning Based:**

1. **Data Collection:** Gather historical incident reports from relevant industries.
2. **Anonymize:** Business critical data.
3. **Data Preprocessing:** Clean and preprocess the data to ensure it is suitable for machine learning analysis.
4. **Model Development:** Develop and train machine learning models using pre-processed data.
5. **Semantic Search and Reclassification:**
6. **Exploration and Reclassification of Cases:** Given the rapidly changing reporting requirements, **zero-shot** and **few-shot** learning methods will be employed to adapt to new classifications quickly.
7. **Translation and Multilingual Embeddings:** Utilizing **translation** and **multilingual embeddings** can enhance the system’s applicability in diverse business environments.
8. **Topic Modelling for Automatic Analyses:**
9. **Clustering Evaluation:** **Dynamic guided topic modelling** will be explored to provide more relevant and business-aligned clusters.
10. **Analyses Based on Clusters:** Various analyses, including **descriptive statistics**, **time series methods**, **early detection**, and **breakpoint detection**, will be conducted based on the identified clusters.
11. **Named Entity Recognition (NER)**
12. **Body Part Extraction:** **Few-shot** learning and **semantic search** will be used to **extract relevant entities**, such as body parts, from incident reports. **Dictionary approaches** will also be considered to enhance accuracy.
13. **Model Evaluation:** **Compare the performance** of the machine learning/LLM model with **manual classification** using metrics such as **accuracy, precision, recall, and F1-score**.
14. **Typical Issues and Solutions:**
15. **Evaluation of Search Results:** **Fine-tuning embeddings**, switching to other **pretrained embeddings**, or creating custom embeddings will be explored to improve search result evaluation.
16. **Reranking Approach:** While a **cross-encoder in SBERT style** may decrease performance, alternative **reranking approaches** will be investigated.
17. **Evaluation of Text Clusters/Topic Modelling:** **Guided topic modelling** with **definable clusters** will be prioritised to align with business needs.
18. **Early Detection of Trends:** **Automatic estimation** methods will be used for early trend detection.
19. **Productive Usage:** Implementing a **delta or batch mode** will be essential to ensure the **system’s practical applicability**.
20. **Implementation and Testing:** Deploy the best-performing model in a real-world setting and monitor its performance over time.
21. **Expected Outcomes:**
22. A significant **reduction in classification errors** compared to manual reporting.
23. Improved **efficiency** in the incident reporting process.
24. Enhanced **decision-making capabilities** due to more accurate incident data.

**Large Language Model Based:**

### **Pipeline 1: Generative Approach (Zero-shot Inferencing)**

**Step 1: Collect and Anonymize Data**

* **Data Collection**: Gather a diverse set of incident reports from various sources.
* **Anonymization**: Remove or mask any sensitive information to ensure privacy and compliance with data protection regulations.

**Step 2: Inferencing**

* **Model Selection**: Use a large language model (LLM) such as LLaMA 3.21B, which can generate text based on the input it receives.
* **Zero-shot Inferencing**: This approach involves using the LLM to generate responses or classifications without any prior training on the specific dataset. The model leverages its pre-existing knowledge to infer the most likely output.

**Step 3: Output**

* **Generated Text**: The LLM produces text that attempts to classify or describe the incident reports based on its understanding. This text can be used for further analysis or as a preliminary classification.

### **Pipeline 2: Classification (LLM with Classification Head)**

**Step 1: Collect and Anonymize Data**

* **Data Collection**: Like Pipeline 1, gather incident reports/data.
* **Anonymization**: Ensure all data is anonymized to protect sensitive information.

**Step 2: LLM with Classification Head (Finetuning)**

* **Model Selection**: Choose an LLM that supports adding a classification head, which is a layer specifically designed for classification tasks.
* **Finetuning**: Train the LLM on the collected dataset, adjusting the model’s parameters to improve its performance on the specific task of incident classification. This involves supervised learning where the model learns from labelled examples.

**Step 3: Output**

* **Classification**: The finetuned model outputs a classification for each incident report, categorizing them into predefined classes based on the training it received.

### **Pipeline 3: Retrieval Augmented Generation (RAG)**

**Step 1: Input**

* **Data Collection**: Collect incident reports as in the previous pipelines.
* **Preprocessing for Class Verification**: Prepare the data to ensure it is suitable for the RAG process, which may involve cleaning and structuring the data.

**Step 2: LLM with RAG**

* **Model Selection**: Use an LLM integrated with a Retrieval Augmented Generation (RAG) framework. RAG combines the strengths of retrieval-based and generative models.
* **Retrieval Component**: The model retrieves relevant documents or data points from a large corpus that can help in generating a more accurate and contextually relevant response.
* **Generation Component**: The LLM then generates text based on the retrieved information, enhancing the quality and relevance of the output.

**Step 3: Classifier (Combined Classes)**

* **Classification**: After generating the text, a classifier is used to categorize the incidents into combined classes. This step ensures that the generated text is aligned with the predefined classification schema.

**Step 4: Output**

* **Class**: The final output is a class label for each incident report, determined by the combined efforts of retrieval, generation, and classification processes.

**Timeline:**

[**https://www.onlinegantt.com/**](https://www.onlinegantt.com/)

Tentative Start Date: 01.12.2024

End Date: 31.05.2025

* Months 0-1: Literature review, data collection, site visit, data pre-processing
* Months 1-3: Model development, Model evaluation and refinement.
* Months 3-5: Implementation and testing in a real-world setting.
* Months 5-6: Analysis of results and thesis writing.

**Conclusion:**

This research aims to address the limitations of manual incident reporting by introducing a machine learning-based classification system. By leveraging historical data, the proposed system is expected to enhance the accuracy and efficiency of incident reporting, ultimately leading to better decision-making and resource allocation.

**Reference:**