**Postdoctoral Research Proposal: Enhancing Industrial Anomaly Detection Through Multimodal Data Integration**

**Phase 1: Benchmarking Existing Anomaly Detection Models on MVTEC-AD**

**Objective:**

* Train and evaluate existing anomaly detection models using the unimodal MVTEC-AD dataset.
* Compare and visualize the performance metrics (accuracy, precision, recall, F1-score, etc.) across different models to establish a baseline.

**Details:**  
We will begin by applying state-of-the-art anomaly detection models to the MVTEC-AD dataset. This phase will involve comprehensive training and evaluation, aiming to understand the strengths and weaknesses of current methodologies. By visualizing statistical outcomes, we will identify performance gaps and areas for potential improvement in anomaly detection tasks within a unimodal context.

**Phase 2: Creating a Dual-Modal Dataset Through Textual Augmentation**

**Objective:**

* Address the lack of multimodal datasets in industrial applications by generating textual descriptions of images in the MVTEC-AD dataset using Large Language Models (LLMs).

**Details:**  
Given the scarcity of industrial multimodal datasets, we propose to create a dual-modal version of the MVTEC-AD dataset. Using advanced LLMs, we will generate detailed textual descriptions that correspond to each image in the dataset. These descriptions will include information about the object, the context, and the nature and location of anomalies. This step effectively transforms MVTEC-AD into a dual-modal dataset, opening new avenues for multimodal analysis.

**Phase 2': Enhancing Textual Representation with Spatial Localization**

**Objective:**

* Improve the precision of textual descriptions by integrating spatial information about the anomalies within the images.

**Details:**  
To increase the accuracy and utility of the generated textual data, we will apply a grid identification technique to the images. This approach will allow the LLMs to generate text that not only describes the anomaly but also specifies its precise location within the image. This spatially aware textual representation will enhance the quality of the dual-modal dataset and provide more granular data for subsequent analysis.

**Phase 3: Multimodal Data Binding and Model Training**

**Objective:**

* Integrate the generated textual data with the original images using vector embedding techniques like ImageBind (Meta) and train the resulting multimodal vectors.

**Details:**  
With the dual-modal dataset prepared, the next step involves binding the image and text modalities into a unified vector representation using advanced techniques like ImageBind. This single-binded vector will serve as the input for anomaly detection models. We will retrain the models from Phase 1 using these multimodal inputs and evaluate the impact of the additional modality on model performance. The goal is to determine whether the integration of text enhances the model's ability to detect anomalies, potentially leading to more robust and accurate detection systems.

**Phase 4: Expanding Modalities for Future Research**

**Objective:**

* Explore the potential of adding further modalities (e.g., heatmaps, depth maps, audio) to the dataset and their impact on anomaly detection.

**Details:**  
If the initial integration of text and images proves successful and unbiased, we will explore the generation of additional modalities, such as heatmaps, depth maps, and even audio signals where applicable. These modalities could provide richer information and context, potentially further enhancing anomaly detection capabilities. This phase will involve experimentation with different combinations of modalities and their effects on detection accuracy, leading to a more comprehensive understanding of multimodal approaches in industrial settings.

**Impact and Future Directions**

This research aims to push the boundaries of anomaly detection in industrial applications by addressing a significant gap in multimodal data availability. By developing and validating a multimodal version of the MVTEC-AD dataset, this project will not only contribute to the academic understanding of multimodal learning but also provide practical tools and insights for industrial applications. If successful, this approach could lead to the creation of more advanced anomaly detection systems that are both more accurate and resilient across various types of industrial data.

**Additional Considerations**

* **Bias and Fairness:** Throughout the research, careful attention will be given to ensuring that the generation of new modalities does not introduce bias, particularly in the textual descriptions. We will employ techniques to evaluate and mitigate potential biases, ensuring that the multimodal data enhances model fairness and accuracy.
* **Scalability:** The proposed methodologies should be designed with scalability in mind, allowing for the easy extension of these techniques to other datasets and industrial applications.