**Testing Plan**

**Purpose Of Testing**

The purpose of this testing plan is to evaluate the reliability, accuracy, and robustness of the proposed AI system designed to predict spacecraft equipment failures. Since real-world testing on actual spacecraft is not feasible in the scope of this project, we will use historical and synthetic telemetry datasets to simulate various operating conditions, including normal performance, degradation patterns, and failure events. Testing ensures that the AI can correctly identify equipment health statuses and respond appropriately to abnormal data.

**Test Scenarios and Cases**

A screenshot of a computer screen

AI-generated content may be incorrect.

**Evaluation Metrics**

We can use the following metrics to evaluate system performance during testing:

* **Accuracy**: Overall correct predictions vs. total predictions.
* **Precision**: Percentage of predicted failures that were actual failures (low false positive rate).
* **Recall (Sensitivity)**: Percentage of actual failures correctly predicted (low false negative rate).
* **F1 Score**: Harmonic mean of precision and recall, balancing both.
* **ROC-AUC Score**: Measures the trade-off between true positive and false positive rates.
* **Confusion Matrix**: Visual representation of prediction categories (True Positive, False Positive, etc.)

**Validation Strategy**

To validate performance, we will split our data into a 70/30 training and testing set and use k-fold cross-validation to ensure generalizability. Additional stress testing will involve simulating harsh conditions like power surges, overheating, or rapid pressure drops. Where real data is lacking, synthetic data will be generated based on documented failure signatures and normal operation baselines. This approach will help ensure the AI model is well-rounded and capable of handling the wide variety of conditions found during space missions.

**Handling Edge Cases**

When it comes to handling edge cases, the AI system must remain stable even when data is missing or contradictory. For example, if temperature data is unavailable or corrupted, the system should fall back on voltage and vibration metrics while logging the issue for review. If multiple sensors give conflicting readings, the model will compute a weighted risk score based on the reliability and priority of each input. If input data is unreadable, the system will escalate the issue with a general alert and defer judgment to mission control or a backup subsystem.

**Conclusion**

In conclusion this testing plan aims to ensure that the proposed predictive AI system can identify potential equipment failures under a wide range of scenarios while maintaining reliability and interpretability. Although real world deployment would involve many additional steps and testing rounds, this plan provides a structured and thoughtful foundation for validating the system’s conceptual design.