Rolls-Royce Aerospace Data Synthesizer

Team Member Names: Varun Senthil, Jai Agarwal, Alejandro Cano, Prisha Rachakonda, Tanay Jain, Mario Guadamillas Jaquete, Georgios Giannakis, Zackary Weinreb, Alex Ye, Vaidehi Guyal, Nicholas Chong, Connor Pfeiffer, Ian Owen Ou, Avyukta Davu, Yohann Viju Kattukaran, Chen Yang, Shravan Suresh, Kan Zhang, Carsten James Schumm, Tyler Kei, Xirui Wang, **Teaching Assistant:** David Alexander Panek

Introduction

Background

- Rolls-Royce requires large volumes of engine sensor data to support research and development
- The Prognostics & Health Management (PHM) team needs realistic synthetic engine data

Problem Statement

- Engineers lack an accessible tool to generate, adjust, and interact with synthetic sensor data

Project Goals

- Need to create a machine learning (ML) model that generates large amounts of realistic, high-quality synthetic time-series engine data
- Deliver an interactive frontend UI so engineers can train models, generate data, and visualize sensor outputs

Methodology

• Data Preparation

Real Rolls-Royce engine sensor data was cleaned and analyzed to understand patterns and relationships between sensors.

• Model Development

Two generative models were used:

1. Diffusion Model: Generates data by gradually removing noise.
2. Transformer Model: Generates sensor values sequentially based on previous values.

• Synthetic Data Generation

Models were trained to produce realistic synthetic engine sensor data.

• Validation

Synthetic data was compared with real data using distribution plots, correlation analysis, and PCA visualization.

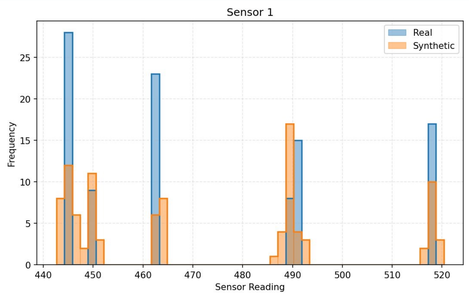
• Interactive Tool

A frontend interface allows engineers to upload data, modify sensors, and explore synthetic datasets.

Workflow

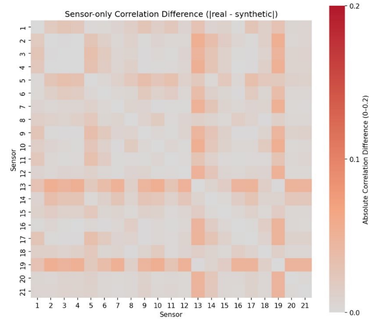
Phase 1: Had four teams' diffusion, transformer, integration, and documentation/validation.
Phase 2: Built on that foundation to implement interpolation, shifting, and propagation so the frontend/UI team could deliver the final user-facing interface.

Graphs

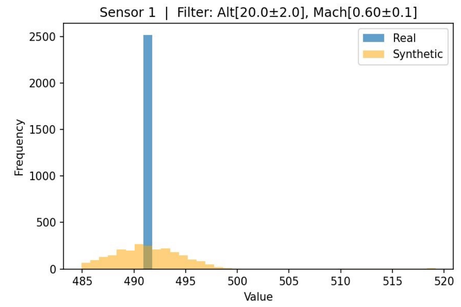


TITLE TEXT

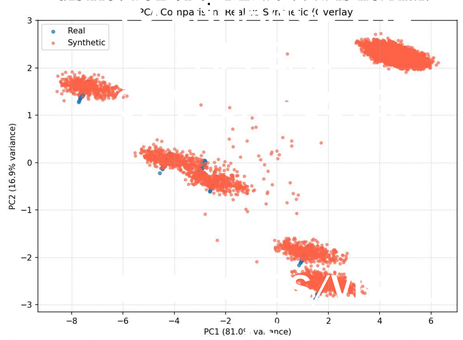
1. Diffusion: Histogram comparing Sensor 1 values from real engine data (blue) and synthetic data (orange). The overlap shows how similar the two distributions are.



Heatmap of the differences between real and synthetic sensor correlations. Lighter colors mean a better match, while darker red shows bigger differences. Most areas are light, indicating the synthetic data preserves sensor relationships well.



2. Transformers: Comparison of real and generated sensor data distributions. When focusing on specific conditions, the distribution of synthetic data is normal.



Scatter plot comparing real and synthetic data, including interpolated values. The synthetic data (orange) clusters in the same regions as the real data (red), with some points in between, showing realistic variation while preserving overall structure.

Impact

- Delivers an interactive tool that empowers Rolls-Royce engineers to generate and manipulate synthetic data
- Allows Rolls-Royce engineers to generate large volumes of realistic engine sensor data on demand, reducing dependency on limited existing datasets
- Accelerates the research and development process by eliminating the time-consuming process of waiting for or collecting real engine test data

Conclusion

• We developed a system that generates realistic synthetic engine sensor data.
• The synthetic data preserves key patterns and relationships from real engine data.
• Engineers can use the tool to experiment with engine scenarios without relying on limited real test data.
• This approach helps accelerate aerospace research and testing.

Future Goals

•Use more real-world variables (engine, environment, flight conditions) to make models more accurate.
•Clean and transform data so it reflects actual engine behavior, not just raw sensor values.
•Make results realistic and easy to use for engineers, including failures and real operating conditions.

Acknowledgements

We would like to thank our mentors **Harry Julijanto** and **Chinmai Trivedi** as well as the **Rolls Royce** team for their guidance and support throughout this project. We also thank our Teaching Assistant David Panek. Lastly, we would like to thank The Data Mine for providing us with this opportunity.