Week 1 Check-in **06/13**

HOUSEHOLD ENERGY CONSUMPTION FORECASTING WITH REAL AND SYNTHETIC DATA

Project Description

This project aims to develop an energy demand forecasting model using real-world and synthetic household electricity consumption data. By leveraging 30-minute interval data across diverse regions and customer profiles, the goal is to accurately forecast energy usage over short-term intervals while keeping consumer privacy under consideration. We will have three settings for training: training with only real data, only synthetic data, and pre-training on synthetic data followed by fine-tuning on real data. We shall compare the resulting models in terms of their accuracy and preservation of privacy. This will support energy providers, policymakers, and researchers in managing grid loads, planning renewable integration, and designing personalized energy solutions.

Dataset Description

1. **TU Delft Electricity Consumption Dataset** Link
   This dataset provides granulated household energy consumption data across Germany, the United Kingdom, Australia, and the Netherlands.
2. **Center for Net Zero (CNZ) Faraday 4.0 Dataset** Link
   A household-level synthetic load profiles dataset generated by a VAE-based model trained on smart meter data from the Octopus Energy Group in the UK. The dataset contains household-level information such as property type and ownership of low-carbon technologies.

Stakeholders

- **Energy Providers**: To optimize their energy supply process.
- **Policy Makers and Regulatory Bodies**: For crafting energy policy and carbon reduction strategies.
- **Smart Technology Developers**: To build adaptive home energy systems based on forecast insights.
- **Researchers & Data Scientists**: For extending forecasting methods or integrating new data sources.
- **Consumers (Indirect Stakeholders)**: Benefit from enhanced personal information privacy, personalized pricing, and improved energy efficiency.

Key Performance Indicators (KPIs)

- **20% improvement in AMSE** over baseline a time-series forecasting model
- **Model comparison**: Demonstrate improvements of generative models compared to the baseline model.
- **Privacy**: Demonstrate the tradeoff between accuracy and privacy when models are trained with synthetic data.

- **Geometric generalization:** Apply models trained with data in the UK to households in several other countries and study their performance.