ACCELERATING GLOBAL ENERGY SYSTEMS RESEARCH WITH OPEN ACCESS TO SYNTHETIC ENERGY DATA
INTRODUCTIONS

GUS CHADNEY
Data Lead

SHENG CHAI
Senior Data Scientist

CENTRE FOR NET ZERO
An impact-driven research unit founded by Octopus Energy
THE DEMAND-SIDE CHALLENGE

DEMAND FLEXIBILITY IS ESSENTIAL
The massive uptake of intermittent renewable energy sources will result in a need of 500 GW of demand flexibility globally by 2030, according to the International Energy Agency.

HOUSEHOLD CONSUMPTION IS KEY
As heat and transport electrify, we need to understand household consumption intimately in order to predict usage and optimise flexibility.

WE NEED SMART METER DATA
Granular smart meter data will unlock pioneering research and innovative data products to plan for electrification and unlock demand flexibility.
ACCESS TO RAW SMART METER DATA IS ESSENTIAL FOR ENERGY RESEARCH

SYNTHETIC DATA ALLEVIATES CONSUMER PRIVACY ISSUES

AN OPEN COMMUNITY FOR SYNTHETIC SMART METER DATA WILL ACCELERATE RESEARCH EFFORTS
CNZ Faraday uses a combination of Variational Autoencoders (VAEs) and Gaussian Mixture Model (GMM) to provide best-in-class synthetic data.

**Cutting-edge techniques**

Faraday was trained on 7 million day profiles over a 1 year period from 20K Octopus Energy UK households.

**Trained on real-world data**

Household profiles can be generated with different LCT mixtures, seasonality and EPC ratings.

**Supports archetypes**
Faraday Alpha V3
About Faraday Alpha V3

The latest version of the Faraday Alpha v3 model for generating synthetic household and smart meter profiles given urban inputs. It works the same as earlier versions: one creates a population of archetype households from synthetic smart meter profiles of that population.

Note however that generating household-level profiles is computationally expensive and there are several limitations in this version:

1. Only the following archetypes are available:
   - F1: Single
   - F2: Single
   - F3: Single
   - F4: Single
   - F5: Single

2. Property Type 3: House with bathroom
3. Location is Toronto, Canada
4. Smart Meter Data is for the month of December, which is included in the data.
5. Profile generation is done at the household level.
6. Maximum number of profiles generated is 1000.
7. Profile generation takes up to 1 hour per row.
8. Feedback on any errors or issues is greatly appreciated.

If you have any questions or feedback, please email us at support@faradayalpha.com.
USE CASES

CURRENT

- TEED Digitisation Project by University of Birmingham
- Better Home Leeds Project by ARUP
- Commercial research projects by industry consultancies such as Parity Projects and Turley
- Other academic research projects by PhDs and Postdocs from University of Manchester and King’s College London

CURRENT POTENTIAL

- Regional, national and global grid “digital twins”
- Future energy system simulations
- Designing smart tariffs
- Greenfield grid design
- Extreme weather resilience planning
- Scenario planning
WHY AN OPEN COMMUNITY?

**STANDARDISATION**
We would like to drive consensus on what “good” looks like for synthetic smart meter data, ensuring quality and privacy.

**COMPETITION**
The performance and ability of the generative algorithms will increase massively if contribution is open.

**VARIABILITY**
Consumption profiles vary globally, multiple contributors will ensure we capture all edge cases for research.

**VOLUME**
Synthetic smart meter data needs to be generated at scale, open-sourcing the algorithms will encourage all holders of real data to do this.
SYNTHETIC DATA ECOSYSTEM

MODEL REPOSITORY

- Standardised APIs / framework to enable:
  - Model training with varied algorithms on arbitrary data sets
  - Evaluation of models to benchmark consistently and ensure quality

- Host algorithm / code for generative models that are vetted against a common evaluation framework

- Community can contribute towards algorithm / evaluation framework as research in the area progresses

DATA REPOSITORY

- Data owners can download algorithm/ code from “Model Repository” to train on their proprietary data to generate synthetic data

- Data owners can donate synthetic data to a Data Repository
**SYNTHETIC DATA ECOSYSTEM**

1. Community members make a Pull Request to the "Model Repository" with their generative model.
2. Apply their algorithm on an open-source dataset e.g. Low Carbon London and produce results.
3. Score the algorithm according to an established evaluation framework and publish results in a benchmark table.
4. Approve/Merge pull requests of model algorithm into main branch.
5. Download the library and choose algorithms that suit their use case to generate synthetic data of their own and donate the data.
6. Download synthetic data for their own use cases/applications.
WHY LF ENERGY?

COMMUNITY
Leverage LF Energy’s expertise creating and growing open-source communities

GOVERNANCE
Build the management framework to ensure quality controls and instil confidence

LICENSING
Navigate and implement correct licenses for usage of software and data

MARKETING
Outbound marketing support to grow community with workshops and events
NEXT STEPS

DEFINITION OF GOOD
We will be publishing a technical paper defining the definition of good that looks at fidelity, utility and privacy metrics.

CONTINUED DEVELOPMENT
We will continue to improve our own generator Faraday, as well as lay the groundwork for the synthetic data ecosystem.

OUTREACH
We will be building up our contact book of interested parties, and plan small, focused workshops.
THANK YOU