# SolarDisagg-LLM: Leveraging Time Series Reprogramming of Large Language Models for Solar Disaggregation

Mozhi Chen, Liang Du
*Villanova University*
Villanova, USA
{mchen08, ldu}@villanova.edu

Shengyi Wang
*University of Arkansas*
Little Rock, USA
swang@ualr.edu

Zhenyu Zhao
*Imperial College London*
London, UK
z.zhao1@imperial.ac.uk

***Abstract*—The rapid proliferation of residential photovoltaic (PV) systems has reshaped electricity demand patterns but obscured behind-the-meter (BTM) generation within net-load measurements. Existing solar disaggregation approaches relying on irradiance proxies or conventional regressors often struggle to generalize across heterogeneous households and weather conditions. This paper presents SolarDisagg-LLM, a reprogramming framework that adapts frozen large language models (LLMs) for time-series solar disaggregation. The proposed model integrates (i) multi-scale patch embeddings for multi-resolution temporal tokenization, (ii) prompt-guidance conditioning to incorporate energy-domain priors, and (iii) scale-wise reconstruction for solar disaggregation. Only lightweight adapters and output heads are trained, ensuring parameter efficiency while preserving the LLM's contextual reasoning capability. Experiments on the real-world data demonstrate that SolarDisagg-LLM achieves the best performance among state-of-the-art baselines in the full-data setting and exhibits strong few-shot generalization across unseen households and weather conditions. Overall, SolarDisagg-LLM establishes a new paradigm for applying LLM reprogramming to energy data analytics, offering a scalable and transferable solution for enhancing residential solar visibility.**

***Index Terms*—Solar Disaggregation, Large Language Model, Time Series Analysis**

## I. Introduction

The global shift toward renewable energy has accelerated the deployment of distributed solar photovoltaic (PV) systems within residential communities. In the first quarter of 2025 alone, the U.S. added 10.8 GW of new solar capacity, with solar accounting for 69% of all new electricity generation capacity added during that period [1]. Residential installations added 1,106 MW in Q1 2025, highlighting the growing role of distributed solar in the electricity system. As rooftop PV penetrations rise, they increasingly shape household consumption profiles and local net loads. While this proliferation supports decarbonization and energy autonomy, it also introduces operational challenges for utilities and aggregators. Because advanced metering infrastructure (AMI) typically records only net demand—that is, the difference between load consumption and behind-the-meter PV generation—the true solar contribution remains hidden. This invisibility complicates key tasks such as short-term load forecasting, demand response management, voltage regulation, and congestion mitigation.

The task of solar disaggregation, i.e., recovering BTM solar generation from net load, has thus become critical for both planning and operational reliability. Traditional approaches often depend on irradiance proxies or regression-based estimators, yet their performance degrades across heterogeneous households and weather conditions [2]. Data-driven methods leveraging high-resolution datasets such as Pecan Street [3] have shown promise, including Bayesian structural time series models and deep neural networks [4]. Nevertheless, three core challenges persist: (i) household load exhibits irregular and stochastic fluctuations that mask PV signatures; (ii) solar profiles follow a diurnal shape yet vary significantly with cloud cover and seasonality; and (iii) generalized, efficient solutions are needed for deployment across homes.

Recent advances in sequence modeling suggest a new opportunity. Large language models (LLMs), pretrained on massive corpora, have demonstrated remarkable capacity to capture long-range dependencies and compositional structures. Beyond natural language, researchers are increasingly exploring their adaptability to other sequential modalities. The key insight is that many sequential datasets share universal structures—tokens, context windows, and dependencies across scales—that LLMs are inherently equipped to model. Importantly, parameter-efficient tuning techniques (e.g., adapters, LoRA, and soft prompts) enable domain reprogramming without retraining billions of parameters, lowering computational cost while preserving generalization. While originally designed for text, LLMs can therefore be *reprogrammed* for time-series tasks by embedding continuous signals into token-like representations [5]. This raises an important question: can we adapt LLMs for energy data, particularly to disentangle solar signals hidden within residential net load?

To answer this, we propose **SolarDisagg-LLM**, a parameter-efficient framework that reprograms frozen large language models (LLMs) for residential solar disaggregation. The framework introduces three key innovations: (i) *multi-scale patch embeddings with vocab alignment*, which tokenize load sequences into overlapping windows across multiple temporal scales and align them with the LLM's vocabulary space to bridge the gap between continuous time-series fea-

979-8-3315-8138-1

tures and discrete language representations; (ii) *domain-aware prompt guidance*, which combines energy-domain priors with learnable soft prompts to steer the LLM toward photovoltaic-specific reasoning patterns; and (iii) *scale-wise projection and reconstruction*, which maps LLM hidden states to per-timestep outputs and applies smoothing to produce realistic solar profiles. Only lightweight adapters and output heads are trained while keeping the LLM backbone frozen, achieving strong adaptability with minimal additional parameters. Comprehensive experiments on the real-world dataset demonstrate that SolarDisagg-LLM consistently outperforms state-of-the-art time-series baselines.

## II. Problem Formulation and Data Description

### A. Problem Formulation

Given a daily net-load sequence $\mathbf{x} \in \mathbb{R}^T$ (e.g., $T$=96 at 15-minute resolution), our objective is to disaggregate the corresponding solar generation $\hat{\mathbf{g}} \in \mathbb{R}^W$ for the same time window. The mapping is defined by a learnable function $f_\theta$ parameterized only by lightweight time series adapters and projection heads, while the LLM backbone remains frozen:

$$\hat{\mathbf{g}} = f_\theta(\mathbf{x}), \quad \mathcal{L} = \|\hat{\mathbf{g}} - \mathbf{g}\|_2^2, \tag{1}$$

where $\mathbf{g}$ denotes the ground-truth solar output. This formulation treats solar disaggregation as a supervised regression task, where the model learns to infer the latent photovoltaic component from aggregated load patterns.

### B. Data Description

A real-world residential energy data from the **Pecan Street** (New York, USA) dataset, which records circuit-level measurements from households equipped with rooftop PV systems, is implemented. All signals represent instantaneous active power (in watts). Residential solar and load profiles exhibit strong diurnal regularities yet complex interactions. Solar generation typically follows a bell-shaped curve, increasing rapidly after sunrise, peaking near midday, and declining toward sunset. In contrast, net load—defined as grid import after subtracting PV generation—is largely governed by human activity, showing pronounced morning and evening peaks. During clear-sky conditions, these two signals exhibit a nearly inverse relationship: when solar output rises, the net load often dips into negative values due to PV backfeed. However, under cloudy or low-irradiance periods, this coupling weakens, and net-load variations become dominated by stochastic household usage rather than solar output. Fig. 1 illustrates three representative daily samples from the Pecan Street dataset, showing both strong and weak coupling scenarios between net load and solar generation. Dashed vertical lines separate each daily segment, with the x-axis repeating 0–24 h per day.

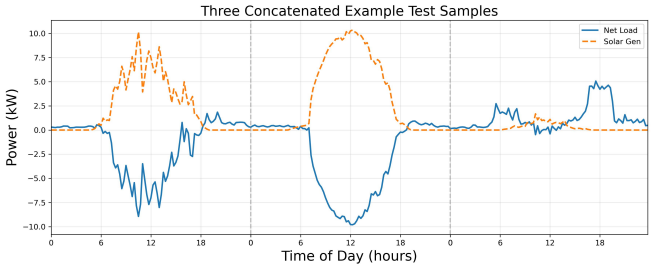


Fig. 1. Three concatenated days' plot from the Pecan Street dataset. Each segment represents one day (0–24 h) of net-load (blue) and solar generation (orange). When solar irradiance is high, the two signals exhibit near-reverse behavior; during cloudy periods, their correlation weakens as stochastic load dominates. Power values are expressed in kilowatts (kW).

## III. Proposed Model

We propose **SolarDisagg-LLM**, a parameter-efficient framework that reprograms frozen large language models (LLMs) for residential solar disaggregation. The overall architecture of the model can be seen in Fig. 2 The model integrates three core components: (1) *multi-scale patch embeddings with vocab alignment*, which tokenize aggregated load sequences across multiple temporal resolutions and align them with the LLM's vocabulary space to bridge the gap between continuous time-series features and discrete linguistic embeddings; (2) *prompt guidance*, which fuses domain-specific textual instructions with learnable soft prompts to guide the LLM toward photovoltaic-related reasoning; and (3) *scale-wise projection and reconstruction*, which maps LLM hidden states to per-timestep solar estimates through a lightweight output head. Only external adapters and projection layers are trained, while the LLM backbone remains frozen, enabling efficient fine-tuning with strong generalization across datasets.

### A. Multi-Scale Patch Embedding and Vocab Alignment

To transform the raw load sequence $\mathbf{x} \in \mathbb{R}^T$ into a representation suitable for LLM processing, multi-scale patch embedding is applied followed by vocab-space alignment.

*1) Multi-Scale Patch Embedding.:* At each temporal scale $s \in \{1, \ldots, S\}$, a one-dimensional convolution extracts overlapping patches with kernel size $p_s$ and stride $\delta_s$:

$$\mathbf{T}_s = \mathrm{Conv1D}(\mathbf{x};\, 1 \to d_p,\, \text{kernel} = p_s,\, \text{stride} = \delta_s) \in \mathbb{R}^{N_s \times d_p}. \tag{2}$$

where $d_p$ is the patch embedding dimension and $N_s$ the number of patches at scale $s$. A learnable embedding $\mathbf{e}_s \in \mathbb{R}^{d_p}$ encodes the temporal resolution of each scale: $\mathbf{T}'_s = \mathbf{T}_s + \mathbf{e}_s$. All scale representations are concatenated to form the complete token sequence:

$$\mathbf{T} = \mathrm{concat}(\mathbf{T}'_1, \mathbf{T}'_2, \ldots, \mathbf{T}'_S) \in \mathbb{R}^{N \times d_p}, \quad N = \sum_s N_s. \tag{3}$$

The combined representation is normalized and projected into the LLM hidden dimension $H$ via a linear transformation:

$$\mathbf{U} = \mathrm{LN}(\mathbf{T})\mathbf{W}_v \in \mathbb{R}^{N \times H}. \tag{4}$$

This hierarchical design efficiently captures both short-term fluctuations and long-term temporal dependencies.

*2) Vocab Alignment:* Following multi-scale patch embedding, the projected time-series tokens $\mathbf{U}$ are aligned with the semantic space of the frozen LLM through a lightweight *Vocab Alignment Adapter*. This module bridges the modality

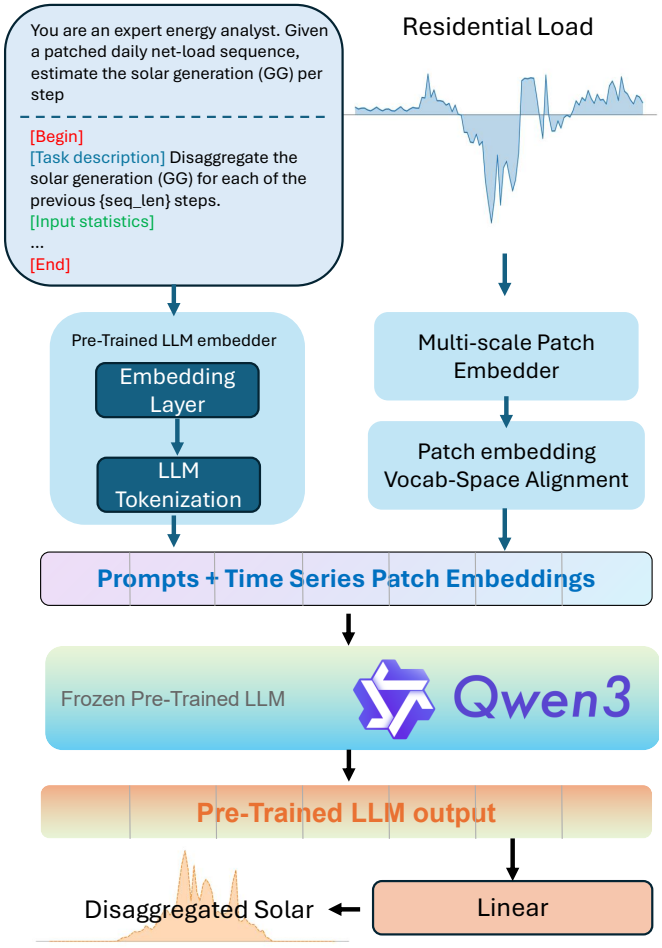


Fig. 2. Overall architecture of the proposed **SolarDisagg-LLM**. The framework combines **multi-scale patch embedding** with **vocab-space alignment** to interface numeric time-series tokens with a frozen pre-trained LLM (`Qwen3` backbone). The textual prompt branch (left) provides domain-aware context via prompt guidance, while the numerical branch (right) transforms multi-resolution load patches into LLM-compatible embeddings. The fused sequence is processed by the frozen LLM and decoded through a lightweight output projection layer for solar disaggregation.

gap between continuous temporal embeddings and discrete language representations. A fixed subset of vocabulary embeddings—referred to as *anchor tokens*—is sampled from the LLM's input embedding matrix to serve as semantic references. As illustrated in Fig. 4, each time-series token interacts with these anchors through a multi-head cross-attention mechanism:

$$\text{head}_i = \text{softmax}\left(\frac{(\mathbf{U}\mathbf{W}_i^Q)(\mathbf{K}\mathbf{W}_i^K)^\top}{\sqrt{H/h}}\right)(\mathbf{V}\mathbf{W}_i^V), \quad (5)$$

$$\mathbf{Z} = \text{LN}\big(\mathbf{U} + \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\mathbf{W}^O\big), \quad (6)$$

where $i = \{1, \ldots, h\}$, $\mathbf{K}, \mathbf{V} \in \mathbb{R}^{K \times H}$ denote the fixed anchor key–value pairs, $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V \in \mathbb{R}^{H \times (H/h)}$ are the projection matrices for the $i$-th head, and $\mathbf{W}^O \in \mathbb{R}^{H \times H}$ is a learnable output projection, respectively. This multi-head design enables each attention head to capture complementary alignments between temporal dynamics and semantic anchors, softly projecting time-series patches toward nearby regions of the LLM embedding manifold.

After obtaining the aligned representation $\mathbf{Z}$, a lightweight linear projection is applied to ensure dimension consistency with the frozen LLM's hidden size: $\mathbf{Z}_{\text{LLM}} = \mathbf{Z}\mathbf{W}_{\text{proj}}$, where $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{H \times D_{\text{LLM}}}$ maps the aligned feature space to the backbone model's hidden dimension $D_{\text{LLM}}$. This projection allows $\mathbf{Z}_{\text{LLM}}$ to seamlessly integrate into the LLM's input pipeline—replacing or augmenting standard token embeddings—so that subsequent transformer blocks can process temporal information within the LLM's latent manifold. Together, the vocab alignment adapter and projection layer complete the modality bridging from time-series patches to semantically grounded, LLM-compatible tokens. The resulting aligned tokens $\mathbf{Z}$ become compatible with the LLM vocabulary space, allowing subsequent transformer layers to process temporal information as coherent language-like tokens.

### B. Prompt Guidance

To infuse domain awareness into the LLM, each input sequence is paired with a structured textual prompt that encodes both task semantics and statistical context. For every batch sample, the model computes basic statistics—including the minimum, maximum, median, and trend direction—along with the top-$k$ autocorrelation lags:

$$\text{stats} = \{\text{min}, \text{max}, \text{median}, \text{trend}, \text{lags}_{1:k}\}.$$

These quantities are inserted into a natural-language template that describes the task to the frozen LLM, for example:

> *"You are an expert energy analyst. Given a patched daily net-load sequence, estimate the solar generation (GG) per step using diurnal patterns but adapt to anomalies."*

The complete prompt is organized into two sections—Task Description and Input Statistics—enclosed by `[Begin]` and `[End]` tokens shown in Fig. 3. This design ensures that both qualitative task intent and quantitative descriptors are embedded in a unified textual form.

After tokenization, the prompt yields a sequence of embeddings $\mathbf{P} \in \mathbb{R}^{L_p \times H}$. We then prepend $K$ learnable soft tokens $\mathbf{S} \in \mathbb{R}^{K \times H}$ to construct the hybrid prefix:

$$\mathbf{Q} = \text{concat}(\mathbf{S}, \mathbf{P}) \in \mathbb{R}^{(K+L_p) \times H}. \quad (7)$$

This prefix merges explicit semantic cues from the textual prompt with trainable soft embeddings, enabling the LLM to capture domain-specific temporal dynamics during downstream solar disaggregation.

### C. Scale-Wise Projection and Reconstruction

After packing and forwarding the concatenated prompt $Q$ and patch embeddings $\mathbf{Z}_i$ through the frozen LLM, we obtain the contextualized output form the combined sequence:

$$\mathbf{O}(i) = \text{concat}(\mathbf{Q}, \mathbf{Z}_i) \in \mathbb{R}^{(K+L_p+N) \times H}. \quad (8)$$

*You are an expert energy analyst. Given a patched daily net-load sequence, estimate the solar generation (GG) per step using diurnal patterns but adapt to anomalies.*

*[Begin]*
*[Task description] Disaggregate the solar generation (GG) for each of the previous {seq_len} steps in the input window;*
*[Input statistics]*
*min value {s['min']:.4f}, max value {s['max']:.4f}, median value {s['median']:.4f}, the trend of input is {s['trend']}, top 5 lags are : {lags_values_str}*
*[End]*

Fig. 3. Example of the constructed **Prompt Guidance** used in **SolarDisagg-LLM**. Each prompt integrates domain knowledge and dataset statistics into a structured format. The sections `[Begin]` and `[End]` enclose the context, consisting of: (1) a **Task Description** specifying the prediction scope (e.g., disaggregate solar generation for the past `{seq_len}` steps), and (2) an **Input Statistics** block summarizing numerical characteristics (min, max, median, trend, and top-5 lags). The resulting prompt is tokenized and embedded by the frozen LLM to provide semantic conditioning for the numeric branch.

Feeding this through the frozen LLM yields contextualized hidden representations:

$$\mathbf{H} = \mathrm{LLM}(\mathbf{O}(i)) \in \mathbb{R}^{(K+L_p+N)\times H}. \tag{9}$$

To isolate the representations corresponding to time-series tokens, we extract the final $N$ positions:

$$\mathbf{H}_{\mathrm{ts}} = \mathbf{H}[-N:,:]. \tag{10}$$

Each hidden state $\mathbf{h}_t \in \mathbb{R}^H$ is then projected to a scalar solar generation estimate via a lightweight regression head:

$$\hat{g}_t = \phi(\mathrm{LN}(\mathbf{h}_t)\mathbf{W}_o + \mathbf{b}_o), \quad t = 1, \ldots, N, \tag{11}$$

where $\phi(\cdot)$ is GELU activation depending on the variant. Collectively, the final solar sequence is generated as:

$$\hat{\mathbf{g}} = [\hat{g}_1, \ldots, \hat{g}_N]^\top \in \mathbb{R}^N. \tag{12}$$

The parameters $\mathbf{W}_o$ and $\mathbf{b}_o$ are learned while the LLM remains frozen, ensuring parameter efficiency. The learning objective is the mean squared error (MSE):

$$\mathcal{L} = \tfrac{1}{W}\sum_{t=1}^{W}(\hat{g}_t - g_t)^2. \tag{13}$$

## IV. Numerical Validation and Results

The proposed framework is evaluated on real-world residential electricity data from the Pecan Street Dataset (NY). The Pecan Street Inc. Dataport provides high-resolution smart meter measurements from households in New York, USA, containing both aggregated net-load and behind-the-meter (BTM) solar generation. This enables the study of solar disaggregation tasks under diverse consumption and generation patterns.

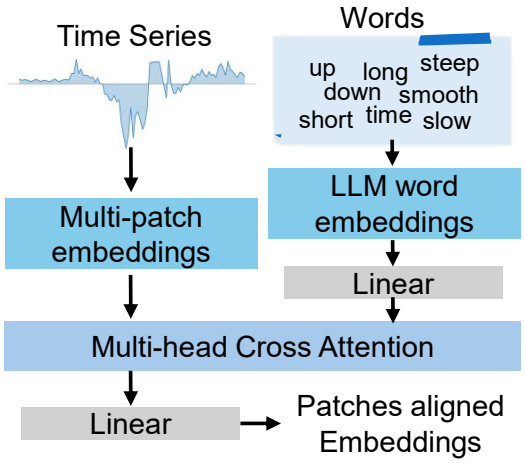


Fig. 4. Overview of the multi-scale patch embedding and vocab alignment.

### A. Baseline Models

To comprehensively evaluate the proposed method, we compare it against a diverse set of state-of-the-art time series models that have demonstrated strong performance in forecasting tasks and are adapted here for solar disaggregation:

- **Informer** [6]: An efficient Transformer variant introducing ProbSparse self-attention to handle long sequences with reduced computational cost.
- **Autoformer** [7]: A decomposition-based Transformer that leverages an auto-correlation mechanism to model long-term temporal dependencies.
- **FEDformer** [8]: A frequency-enhanced decomposition Transformer that combines time- and frequency-domain representations for improved learning efficiency.
- **DLinear** [9]: A lightweight linear model that decomposes time series into trend and seasonal components for efficient and interpretable forecasting.
- **PatchTST** [10]: A Transformer-based model that employs a patching strategy to capture local temporal patterns and long-range dependencies.
- **LightTS** [11]: A simplified Transformer architecture that improves efficiency through lightweight channel mixing and selective token interactions.

### B. Implementation Details

We adopt multi-scale patching with scales $\{(8,4),(24,12),(48,24)\}$ for an input length of $T$=96, capturing intra-hour, hourly, and multi-hour patterns. Each patch is projected to a $d_p$=1024-dimensional embedding. The prompt prefix consists of $K$=16 learnable soft tokens combined with textual instructions. We use Qwen3-0.6B [12] as the frozen backbone LLM. Training is performed for 300 epochs using AdamW ($\eta$=$5\times10^{-4}$) with cosine decay, warmup, and gradient clipping of 1.0.

### C. Model Performance

Table I compares the proposed **SolarDisagg-LLM** with several state-of-the-art time-series baselines on the Pecan

TABLE I
PERFORMANCE COMPARISON ON THE PECAN STREET DATASET. A LOWER MSE, RMSE OR MAE INDICATES A BETTER PREDICTION

| Model | MSE | RMSE | MAE |
|---|---|---|---|
| Informer | 0.5311 | 0.7288 | 0.2397 |
| Autoformer | 2.9009 | 1.7032 | 1.0106 |
| FEDformer | 0.5000 | 0.7071 | 0.2674 |
| DLinear | 0.6532 | 0.8082 | 0.3331 |
| PatchTST | 0.5291 | 0.7274 | 0.2941 |
| LightTS | 0.5396 | 0.7346 | 0.3337 |
| **Proposed** | **0.4098** | **0.6401** | **0.2646** |

TABLE II
FEW-SHOT LEARNING PERFORMANCE ON THE PECAN STREET DATASET. A LOWER MSE, RMSE OR MAE INDICATES A BETTER PREDICTION

| Model | MSE | RMSE | MAE |
|---|---|---|---|
| Informer | 0.6403 | 0.8002 | **0.3270** |
| Autoformer | 2.8900 | 1.7000 | 1.0914 |
| FEDformer | 0.5846 | 0.7646 | 0.3309 |
| DLinear | 0.7200 | 0.8486 | 0.3806 |
| PatchTST | 0.6513 | 0.8070 | 0.3647 |
| LightTS | 0.6733 | 0.8205 | 0.3970 |
| **Proposed (Few-shot)** | **0.5751** | **0.7583** | 0.3778 |

Street dataset. Despite using a compact **Qwen3-0.6B** backbone with a total of 604.6M parameters—of which only 1.42% (8.59M) are trainable when the LLM is frozen—our model achieves the lowest reconstruction error across all metrics (MSE = 0.4098, RMSE = 0.6401, MAE = 0.2646), surpassing all SOTA models while requiring far fewer trainable parameters and computational cost than fully fine-tuning the LLM. This demonstrates that a small, efficiently tuned subset of parameters can effectively adapt the pretrained backbone to the solar domain through multi-scale patch embedding, vocab alignment, and prompt-guided adaptation. These results highlight that lightweight LLM reprogramming can deliver state-of-the-art performance in solar disaggregation with strong generalization capability and excellent training efficiency.

### *D. Few-shot Learning*

To address data scarcity and enhance generalization to unseen residents, we conduct a few-shot learning experiment using only 10% of the training data, 10% for validation, and the rest 80% as test set. As shown in Table II, our model achieves the lowest MSE and RMSE among all baselines, demonstrating a strong ability to transfer knowledge across households with minimal supervision. Even when trained on a small subset of residents, the model preserves stable reconstruction quality for unseen ones, validating the effectiveness of prompt-guided representation learning and multi-scale temporal embeddings in capturing shared load–solar patterns. Compared to the state-of-the-practice solar disaggregation algorithms utilized by system operators [13] or at household level [14], the proposed **SolarDisagg-LLM** is not only data-efficient but also highly generalizable, enabling reliable solar disaggregation across diverse residential settings with limited labeled data.

## V. CONCLUSIONS

This work introduced **SolarDisagg-LLM**, a novel framework that reprograms frozen large language models for residential solar disaggregation. By unifying multi-scale patch embeddings, vocab-space alignment, and prompt-guided adaptation, the model effectively bridges numeric time-series inputs with language-based reasoning. Comprehensive evaluations on the Pecan Street dataset demonstrate that SolarDisagg-LLM consistently outperforms state-of-the-art Time Series baselines across both full-data and few-shot settings. The few-shot results, in particular, reveal the model's strong adaptability when trained on limited data, while maintaining low reconstruction errors. Future work will explore (i) larger backbone models for enhanced temporal reasoning, (ii) multi-modal conditioning using irradiance and weather data, and (iii) transfer learning across heterogeneous grid regions. Overall, this study highlights the potential of LLM reprogramming as a generalizable, parameter-efficient solution for time-series understanding in power systems. Future work will explore incorporating multi-modal conditioning and physics-informed priors into the LLM reprogramming process to improve interpretability with power system knowledge.

## REFERENCES


[1] Solar Energy Industries Association (SEIA) and Wood Mackenzie, "Solar market insight report q2 2025," https://www.seia.org/research-resources/solar-market-insight-report-q2-2025, 2025.

[2] S. Ziyabari, L. Du, and S. K. Biswas, "Multibranch attentive gated resnet for short-term spatio-temporal solar irradiance forecasting," *IEEE Transactions on Industry Applications*, vol. 58, no. 1, pp. 28–38, 2022.

[3] Pecan Street Inc., "Dataport database," https://dataport.pecanstreet.org, accessed September 2025.

[4] S. Wang, L. Du, and X. Chen, "Bayesian active learning-based soft data space calibration for system-wise aggregate flexibility characterization," *IEEE Transactions on Smart Grid*, vol. 16, no. 4, pp. 3017–3029, 2025.

[5] M. Jin *et al.*, "Time-LLM: Time series forecasting by reprogramming large language models," in *International Conference on Learning Representations (ICLR)*, 2024.

[6] H. Zhou *et al.*, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *The Thirty-Fifth AAAI*, vol. 35, no. 12. AAAI Press, 2021, pp. 11 106–11 115.

[7] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021, pp. 22 419–22 430.

[8] H. Zhou *et al.*, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *Proceedings of the 39th International Conf. Machine Learning (ICML)*, 2022, pp. 27 268–27 286.

[9] A. Zeng, M. Chen, L. Zhang, Q. Xu, and D. Li, "Are transformers effective for time series forecasting? a simple DLinear baseline," *arXiv preprint arXiv:2205.13504*, 2023.

[10] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," in *International Conference on Learning Representations*, 2023.

[11] D. Campos *et al.*, "Lightts: Lightweight time series classification with adaptive ensemble distillation," *Proceedings of the ACM on Management of Data*, vol. 1, no. 2, p. 1–27, Jun. 2023. [Online]. Available: http://dx.doi.org/10.1145/3589316

[12] A. Yang *et al.*, "Qwen3 technical report," 2025. [Online]. Available: https://arxiv.org/abs/2505.09388

[13] D. Moscovitz, Z. Zhao, L. Du, and X. Fan, "Semi-supervised, non-intrusive disaggregation of nodal load profiles with significant behind-the-meter solar generation," *IEEE Transactions on Power Systems*, vol. 39, no. 3, pp. 4852–4864, 2024.

[14] L. Du *et al.*, "Self-organizing classification and identification of miscellaneous electric loads," in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1–6.