Llama 3.1 and Mistral 7B Robustness Under Distribution Shifts in Energy Forecasting

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How do Llama 3.1 and Mistral 7B differ in robustness scores against distribution shifts when transferring from synthetic battery datasets to real-world renewable energy forecasting tasks. 17 claims were extracted from source literature; 4 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Lag-Llama: Towards Foundation Models for Probabilistic Time Series Forecasting. Research question: How do Llama 3.1 and Mistral 7B differ in robustness scores against distribution shifts when transferring from synthetic battery datasets to real-world renewable energy forecasting tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.4/10.

3 Results

13 papers retrieved. 17 claims extracted; 4 independently verified. Quality review score: 5.4/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Lag-Llama demonstrates state-of-the-art performance across diverse datasets from different domains after finetuning.	✓	0.18
Lag-Llama emerges as the best general-purpose model without any knowledge of downstream datasets.	✓	0.21
Lag-Llama shows strong few-shot adaptation performance on previously unseen datasets, across varying fractions of data h	✓	0.16
The diversity of the pretraining corpus used to train Lag-Llama is investigated, and scaling laws of Lag-Llama with resp	×	0.08
Statistical models like ARIMA, ETS, and Theta models have been the cornerstone of time series forecasting for decades.	×	0.09
ARIMA uses autocorrelation to forecast future values.	×	0.03
ETS models decompose a time series into its fundamental components, allowing for more nuanced forecasting that captures	×	0.05
Theta models apply a decomposition technique combining both long-term trend and seasonality.	×	0.01
Statistical models assume linear relationships and stationarity in time series data, which is often not the case in real	×	0.07
Statistical models may require extensive manual tuning and domain knowledge to select appropriate models and parameters	×	0.07
Neural forecasting is a rapidly developing research area following the explosion of machine learning.	×	0.06
Various architectures have been developed for neural forecasting, starting with RNN-based and LSTM-based models.	×	0.04
Transformers have been proposed for time series forecasting, processing the input time series in different ways to be di	×	0.11
Alternative strategies to vanilla attention have been proposed to build better models tailored for time series.	×	0.09
The main goal of Lag-Llama is to apply the foundation model approach to time series data and investigate the extent of t	✓	0.19
Lag-Llama considers a dataset of D $\geq$1 univariate time series, sampled at a specific discrete set of time points.	×	0.10
The univariate probabilistic time series forecasting problem involves modelling an unknown joint distribution of the P f	×	0.13

References

- http://arxiv.org/abs/2306.13552v2
- http://arxiv.org/abs/2310.08278v3
- http://arxiv.org/abs/2206.03669v3