servicenow

# Overcoming the Modality Gap in Context-Aided Forecasting

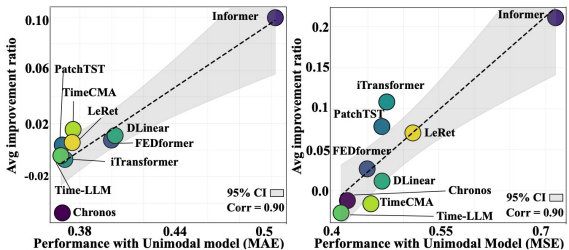
CAF-7M

V. Zheng[1,2] É. Marcotte[1] A. Ashok[1,3,4] A. R. Williams[1,3,4] L. Sun[2]
A. Drouin[1,4] V. Zantedeschi[1]

[1]ServiceNow AI Research [2]McGill [3]UdeM [4]Mila

TSALM Workshop, ICLR 2026

# The Modality Gap in Time Series Forecasting




Zhang et al., 2025 — *When Does Multimodality Lead to Better Time Series Forecasting?*

Surprising finding:

- Context improvement correlates with unimodal error (Corr = 0.90)
- Context helps *weaker* models most
- Best unimodal models beat multimodal ones

⇒ Context doesn't help strong models. Why?

# Our Hypothesis

## The problem is the data, not the models

Existing CAF datasets have **poor context quality** — text is scarce and never verified to be *complementary*.

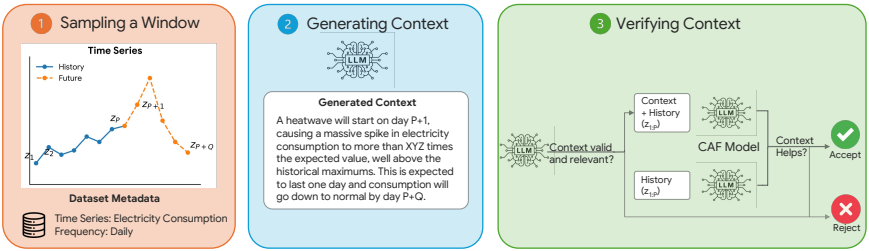
Without verification, we can't distinguish:

- model *can't use* context vs. context *isn't useful*

## Our approach

A semi-synthetic pipeline generating **verifiably useful** contexts at scale ⇒ **CAF-7M** (7M windows).

# CAF-7M: Data Pipeline




1. **Window Sampling:** 65 datasets, 11 domains
2. **Context Generation:** `Llama-3.3-70B` generates scenarios from history + future
3. **Verification:** Two-stage quality filter (next slide)

# Context Verification: Ensuring Quality

LLM-generated contexts can *appear* informative but add nothing useful.
Two-stage filter:

## Stage 1: Plausibility Check

`GPT-5.2` checks: does the context mention the **target variable** and describe a **plausible scenario**?

## Stage 2: Complementarity Check (CRPS-based)

A **Direct Prompt** forecaster (`GPT-5.2`) predicts *with* and *without* context. Keep only windows where context **lowers the CRPS**.

Filtering applied *only to the test set* (**904** verified windows).
Training data (**7,433,239** windows, 40 datasets) is unfiltered for cost reasons.

# CAF-7M

Training on *unfiltered* data ⇒ **clear gains** on the verified test set.


Demo

vincent-zheng.com/DoubleCast

huggingface.co/datasets/ServiceNow/CAF_7M  github.com/ServiceNow/DoubleCast

arxiv.org/abs/2603.12451