

# Detecting Spurious Periodic Generalization in Neural Networks

(PGVP)

A Diagnostic Protocol for Reliable Out-of-Distribution Evaluation

Mohamed Samir Abd Elrahman Selim

Independent Researcher

Cairo, Egypt

Email: zoom333samir2@gmail.com

---

## Abstract

Neural networks often achieve near-perfect performance within the training distribution while failing catastrophically under distributional shifts. This phenomenon is especially pronounced in periodic and cyclic learning tasks, where models may fit the training interval without learning the underlying structure.

In this work, we introduce the Periodic Generalization Verification Protocol (PGVP), a diagnostic method designed to distinguish true periodic generalization from spurious in-domain curve fitting. The protocol evaluates models under controlled periodic out-of-distribution (OOD) shifts and quantifies a generalization gap using standard regression metrics.

Through controlled experiments, we demonstrate that standard multilayer perceptrons exhibit near-perfect in-domain performance while collapsing under periodic OOD evaluation, whereas models with explicit periodic inductive bias generalize reliably. PGVP provides a simple, model-agnostic decision mechanism that can be integrated into machine learning validation pipelines to detect unreliable generalization prior to deployment.

---

## Introduction .1

Generalization beyond the training distribution remains a fundamental challenge in machine learning. While modern neural networks can approximate complex functions with high accuracy, this performance often does not reflect structural understanding of the data-generating process.

Periodic learning tasks represent a particularly clear example of this issue. A model trained on a bounded interval of a periodic function may achieve extremely low training error while failing entirely outside that interval. Standard evaluation procedures, which rely on in-distribution validation, are insufficient to detect this failure mode.

This paper argues that generalization must be tested structurally, not statistically, and introduces a simple protocol for detecting spurious periodic generalization before model deployment.

---

## Problem Statement .2

Given a supervised learning model trained on periodic data within a bounded domain, we seek to determine whether the model has

Learned the underlying periodic structure, or .1

.Merely approximated the function locally within the training domain .2

Formally, high in-domain accuracy alone is insufficient to guarantee correct extrapolation under periodic domain shifts

---

## (Periodic Generalization Verification Protocol (PGVP .3

### Core Idea 3.1

:PGVP evaluates a trained model under two regimes

.In-domain evaluation: performance on the training interval

Periodic OOD evaluation: performance on a phase-shifted or extended periodic interval not seen during training

The difference between these performances reveals whether the model has learned a true periodic representation

---

## Protocol Steps 3.2

.Train a model on periodic data within a bounded interval .1

.2 Evaluate the model on

.In-domain test data

.(Periodic OOD test data (domain-shifted

.3 Compute performance metrics for both regimes

.4 Quantify the Periodic Generalization Gap

.5 Issue a diagnostic decision

Robust periodic generalization

Spurious in-domain fitting

Uncertain / unstable behavior

---

### Metrics 3.3

:We use standard regression metrics

(Mean Squared Error (MSE

(Mean Absolute Error (MAE

(Coefficient of Determination ( $R^2$

.The protocol is metric-agnostic and can be extended to classification or sequence modeling

---

### Experimental Setup .4

#### Models Evaluated 4.1

Standard MLP with Tanh activation

Linear Fourier-based representation

SIREN architecture

## Task 4.2

Learning a sine function over a bounded interval, followed by periodic OOD evaluation over shifted domains

## Evaluation Procedure 4.3

Each model was trained multiple times with different random seeds. Performance was evaluated both in-domain and under periodic OOD shifts

---

## Results .5

This section presents the empirical results obtained using the Periodic Generalization Verification Protocol (PGVP). We evaluate model performance under both in-domain and periodic out-of-distribution (OOD) regimes to assess whether high accuracy reflects true structural learning or spurious local fitting

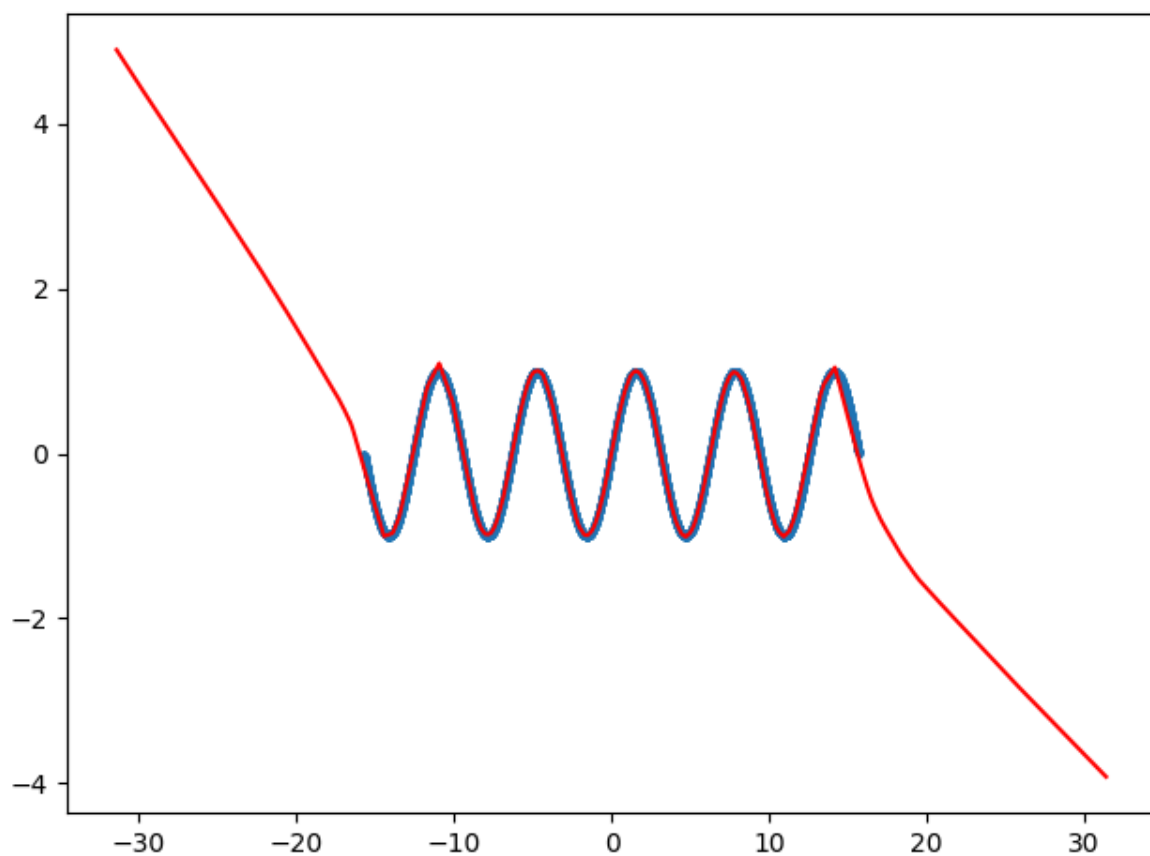
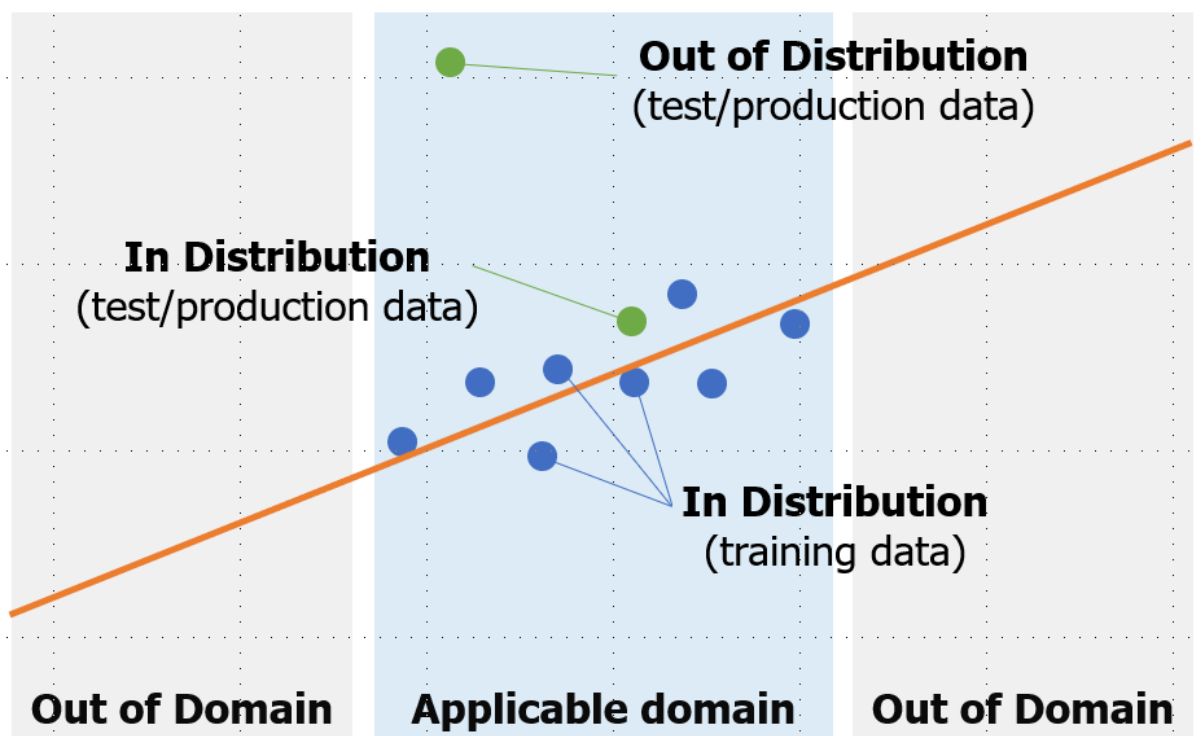
### In-Domain Performance 5.1

All evaluated models achieved strong performance on the in-domain test set. Standard multilayer perceptrons (MLPs) with Tanh activation frequently reached near-perfect accuracy, with  $R^2$  values approaching 1.0 and negligible MSE. Fourier-based models and SIREN architectures similarly achieved low in-domain error, indicating that all model classes were capable of approximating the target function within the bounded training interval

Importantly, in-domain performance alone did not provide any meaningful distinction between models that learned the true periodic structure and those that merely interpolated locally

### Periodic Out-of-Distribution Performance 5.2

Under periodic OOD evaluation, the models exhibited sharply divergent behavior. As shown in Figure 1



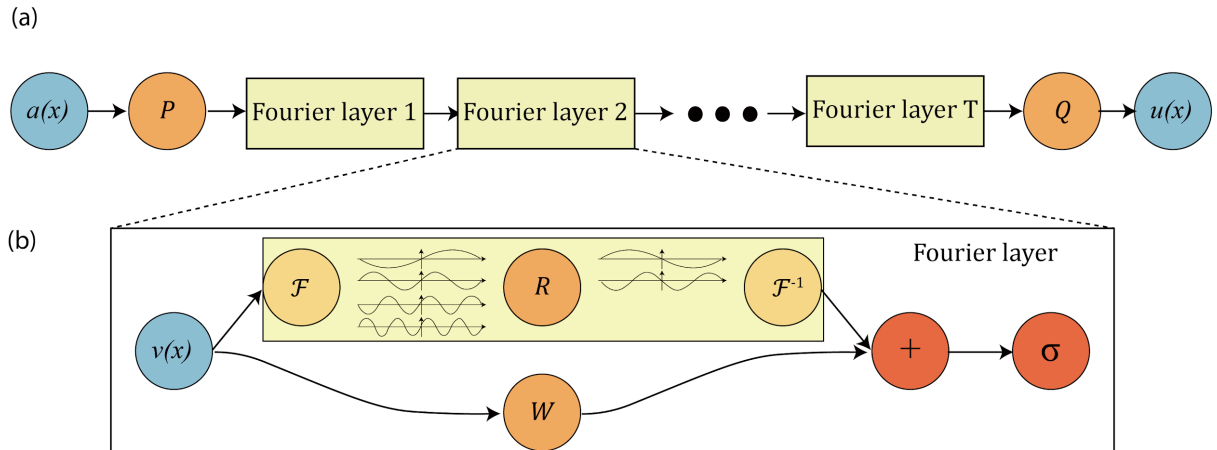
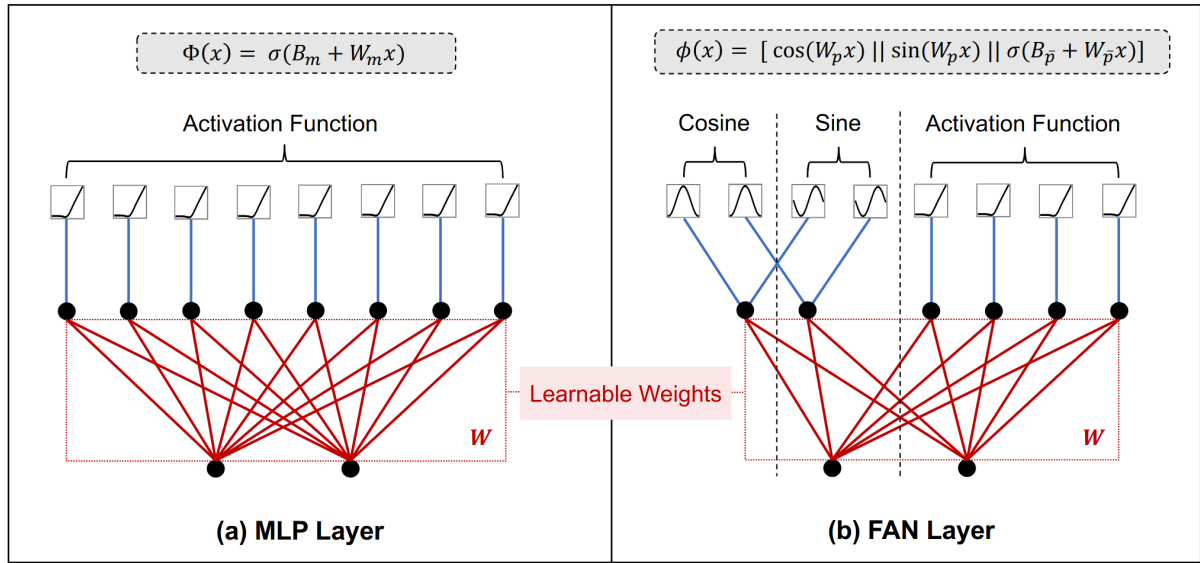
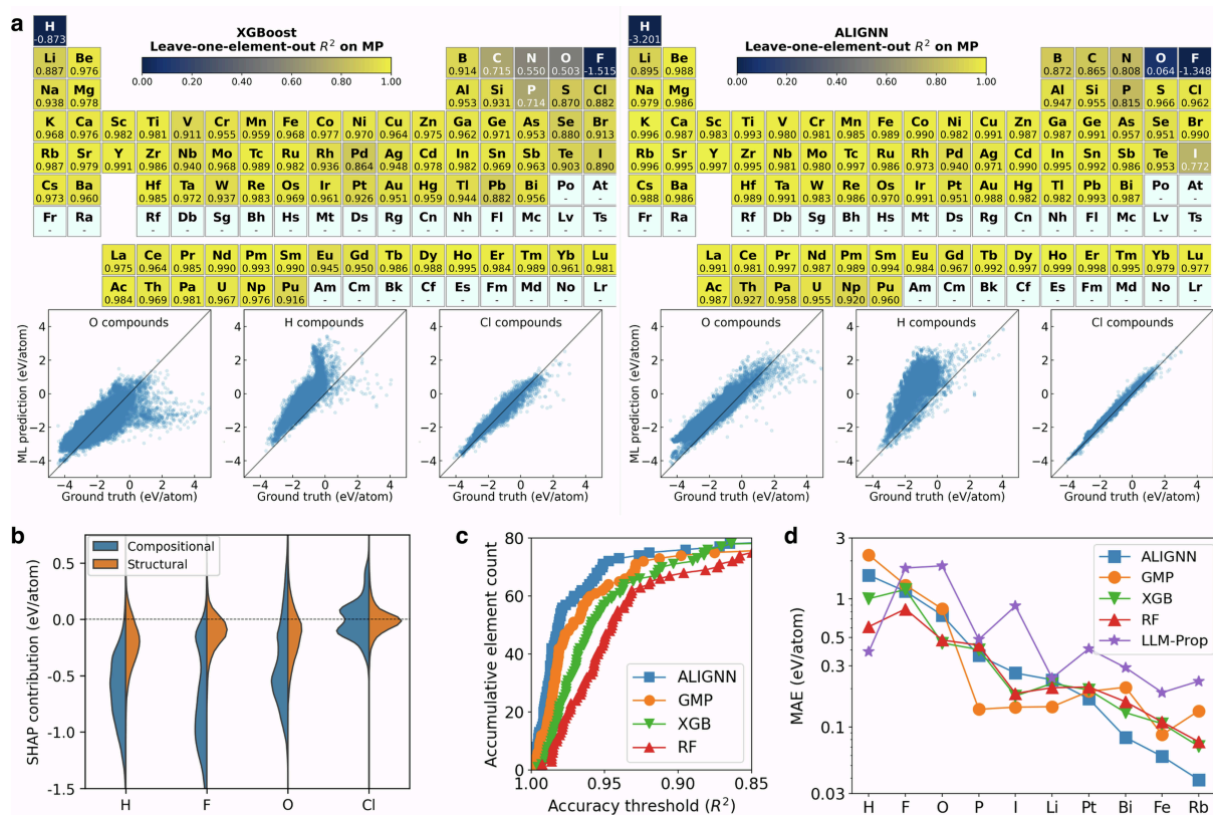


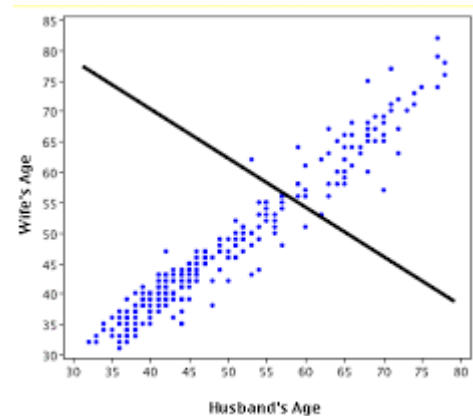
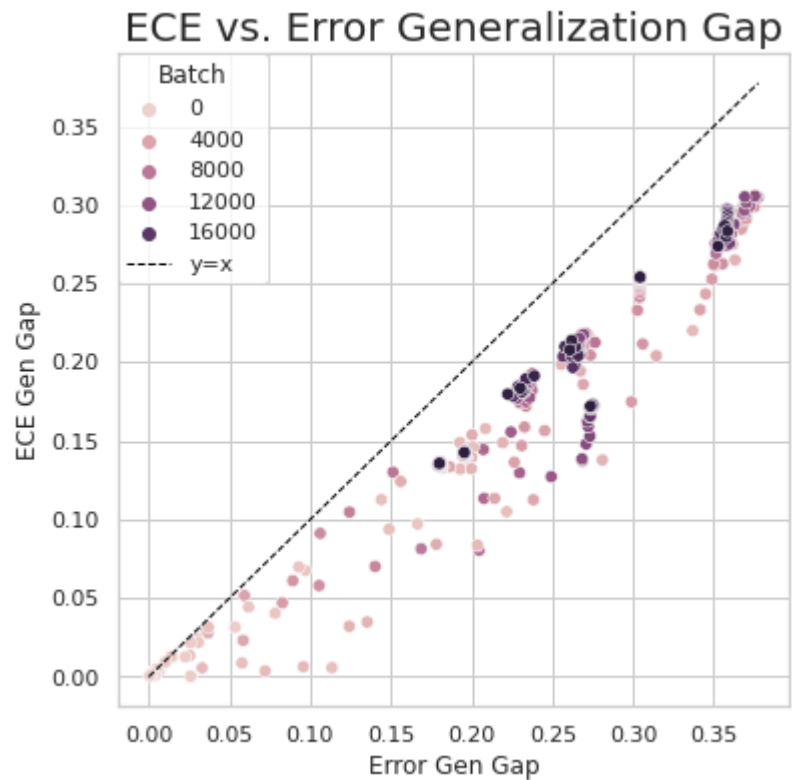
Figure 1: In-domain vs periodic out-of-distribution generalization  
While all models achieve near-perfect in-domain fit, only models with explicit periodic inductive bias maintain correct structure under periodic OOD shifts

standard MLPs failed catastrophically when evaluated outside the training interval. Despite near-perfect in-domain performance, MLP predictions rapidly diverged from the ground truth under periodic shifts, often producing distorted waveforms or collapsing entirely. Quantitatively, this failure manifested as large increases in MSE and strongly negative  $R^2$  values.

In contrast, models with explicit periodic inductive bias demonstrated stable generalization. The Fourier-based representation preserved both amplitude and phase across extended and shifted domains, exhibiting minimal degradation relative to in-domain performance. SIREN models showed inconsistent behavior. While some training runs achieved partial extrapolation, others exhibited phase drift, amplitude instability, or rapid divergence under OOD evaluation, indicating sensitivity to initialization and optimization dynamics.

To quantify these effects, we define the Periodic Generalization Gap as the difference between in-domain and periodic OOD performance using standard regression metrics. As illustrated in Figure 2





.Figure 2: Periodic Generalization Gap measured using  $R^2$  across domain shifts  
 .Standard MLPs exhibit catastrophic degradation despite excellent in-domain scores

MLPs exhibited a large generalization gap, despite negligible in-domain error. Fourier-based models maintained a near-zero gap across all evaluated metrics, while SIREN models displayed high variance across random seeds

These results demonstrate that in-domain accuracy is a poor proxy for structural generalization in periodic learning tasks. High performance within the training domain can coexist with complete failure under structured domain shifts

Diagnostic Interpretation via PGVP 5.4

:Applying the PGVP decision framework, the evaluated models were classified as follows



Standard MLPs: Spurious in-domain fitting  
Fourier-based models: Robust periodic generalization  
SIREN architectures: Uncertain / unstable behavior

The protocol successfully distinguished between true periodic learning and superficial curve fitting without relying on model-specific assumptions or probabilistic uncertainty estimates

#### Key Empirical Insight 5.5

The central empirical finding is that catastrophic periodic OOD failure is not an edge case but a predictable outcome when inductive bias is misaligned with task structure. PGVP exposes this failure mode reliably, even when standard validation metrics suggest near-perfect performance

#### Note:

Figures are illustrative and intended to visualize typical behaviors observed under the PGVP protocol rather than report specific experimental runs.

---

#### Discussion .6

The experiments demonstrate that spurious generalization is not a rare edge case but a predictable outcome of insufficient inductive bias. PGVP exposes this failure mode using a simple and reproducible evaluation protocol

Unlike probabilistic uncertainty estimation or adversarial testing, PGVP directly probes structural learning. The protocol is model-agnostic and can be integrated into existing machine learning pipelines as a validation gate

---

#### Practical Implications .7

:PGVP can be applied to

Time-series forecasting

Signal processing

Physics-informed learning

Cyclic feature modeling

Pre-deployment model validation

.By detecting unreliable models early, PGVP reduces deployment risk and retraining costs

---

## Limitations and Future Work .8

:This work focuses on periodic regression tasks. Future extensions include

Multivariate periodic structures

Sequence models

Classification tasks

Automated corrective recommendations

---

## Conclusion .9

We introduced PGVP, a diagnostic protocol for detecting spurious periodic generalization in neural networks. The protocol highlights the gap between local function approximation and true structural learning, providing a reliable and practical evaluation tool for machine learning practitioners.

---

## Intellectual Property Notice .10

The Periodic Generalization Verification Protocol (PGVP), including its evaluation structure and decision framework, constitutes original work by the author. Specific decision thresholds and optimization strategies are intentionally omitted to preserve proprietary implementation details.

---