This Time is Different: An Observability Perspective on Time Series Foundation Models

Ben Cohen[*1 2] **Emaad Khwaja**[*1 2]
Youssef Doubli[1 2] Salahidine Lemaachi[1 2] Chris Lettieri[1 2] Charles Masson[1 2] Hugo Miccinilli[1 2] Elise Rame´[1 2] Qiqi Ren[1 2]
Afshin Rostamizadeh[1 2] Jean Ogier du Terrail[1 2] Anna-Monica Toon[1 2] Kan Wang[1 2] Stephan Xie[1 2 3] Zongzhe Xu[1 2] Viktoriya Zhukova[1 2]
David Asker[2] Ameet Talwalkar[2 3] Othmane Abou-Amal[2]

NEURAL INFORMATION PROCESSING SYSTEMS

[*]Equal contribution, [1]Core Contributor (listed alphabetically), [2]Datadog AI Research, [3]Carnegie Mellon University
Correspondence: {ben.cohen, emaad, ameet, othmane}@datadoghq.com

Motivation

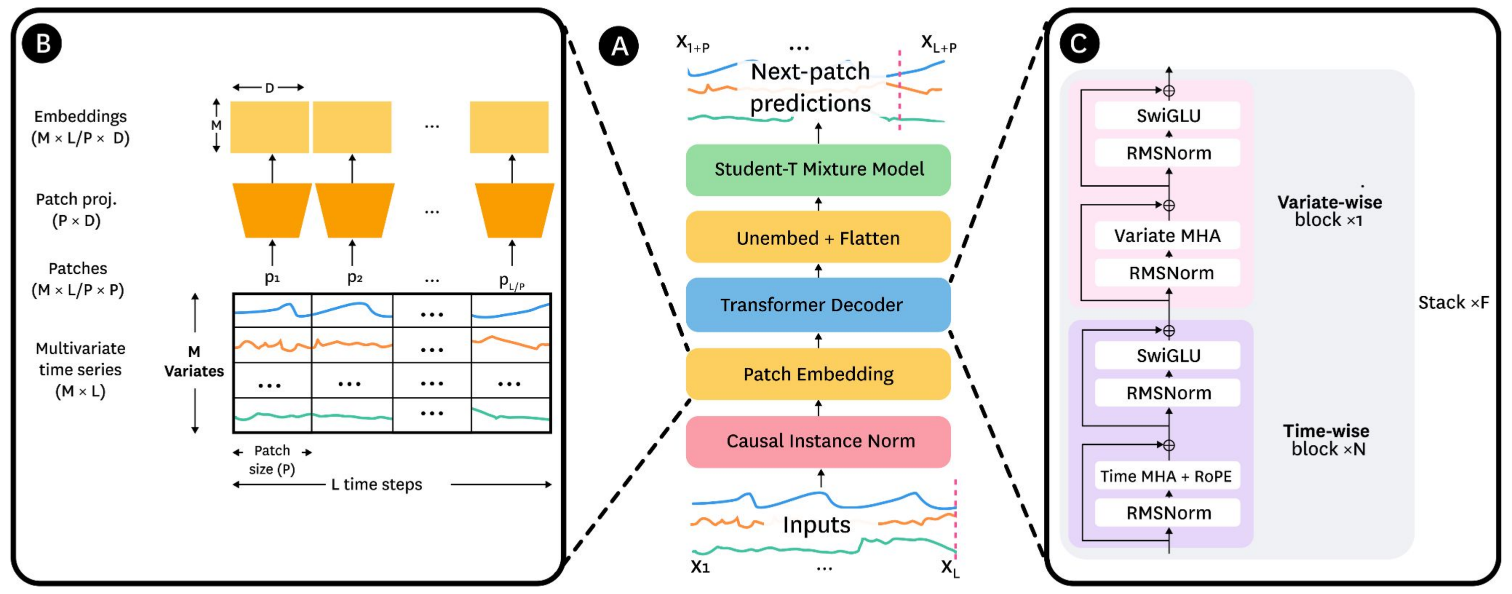
- **Observability time series are different:** Metrics are high-volume, multivariate, heavy-tailed, and nonstationary compared to standard time-series benchmarks.
- **Gap in models and data:** Existing TSFMs underperform on observability, and no dedicated benchmark or model existed before Toto and BOOM.

Contributions

- **Toto:** A pretrained, zero-shot time-series foundation model achieving state-of-the-art performance on BOOM, GIFT-Eval, and LSF. The first TSFM optimized for observability.
- **BOOM:** large observability benchmark: Massive, high-dimensional real telemetry benchmark released with Toto weights and full evaluation code.

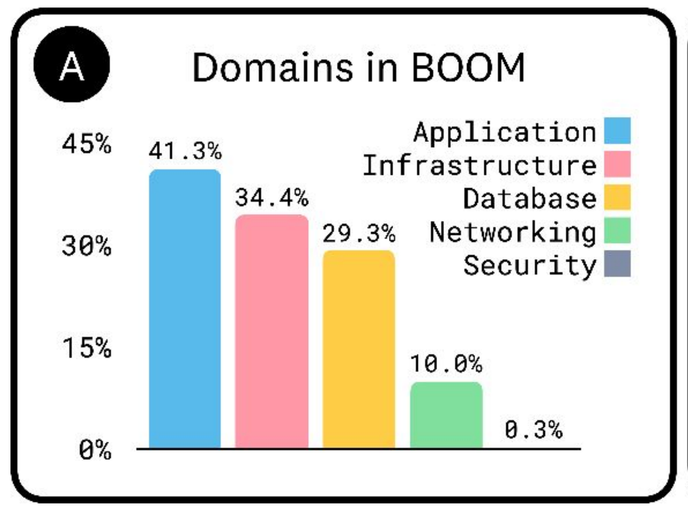
Toto Architecture

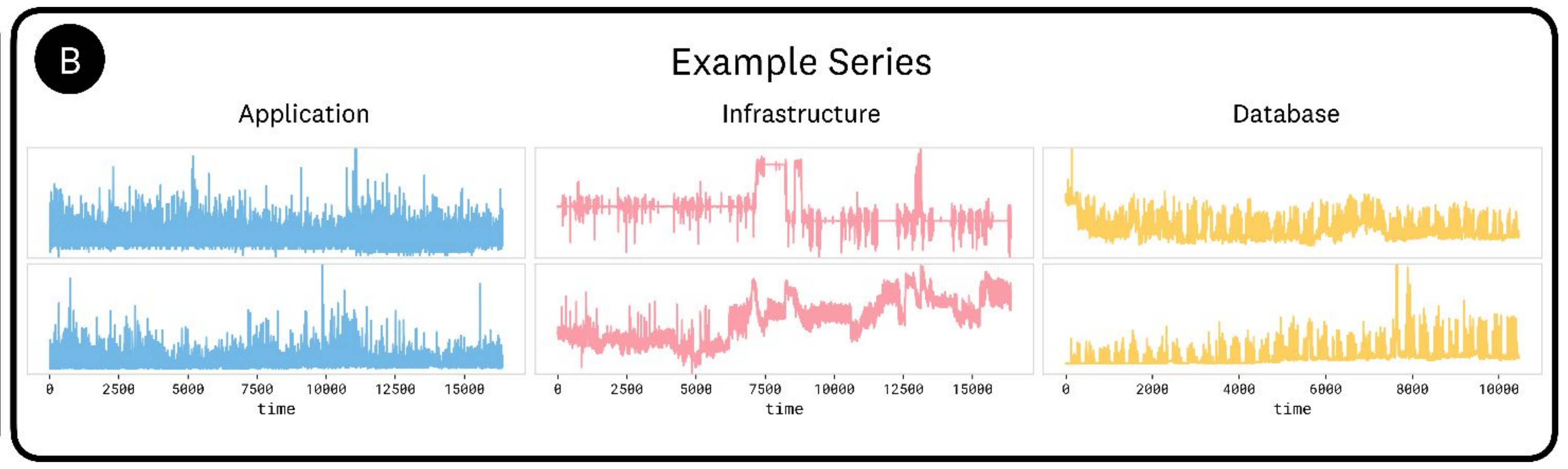
- Ⓐ **Decoder-only transformer:** Next-patch prediction for zero-shot multivariate forecasting.
- **Causal scaling:** Patch-based causal instance normalization to handle strong nonstationarity.
- Ⓑ **Patch embeddings:** Non-overlapping patches compress long contexts into manageable token sequences.
- Ⓒ **Factorized attention:** Alternating time-wise and variate-wise blocks to scale to high-cardinality series.
- **Student-T mixture head:** Heavy-tailed probabilistic forecasts robust to spikes and skew.



- **Composite loss:** NLL plus Cauchy loss for stability on outliers.

BOOM Dataset





- Ⓐ **Real observability telemetry:** 350M points across 2,807 metric queries from Datadog's internal staging environment, disjoint from Toto's pretraining data.
- **Rich coverage:** Application, infrastructure, database, networking, and security metrics spanning gauges, rates, counts, and distributions.
- Ⓑ **Hard, high-dimensional stats:** Median ≈60 variates per series with more spikes, nonstationarity, and heavy tails than standard time-series benchmarks.

Results

GIFT-Eval

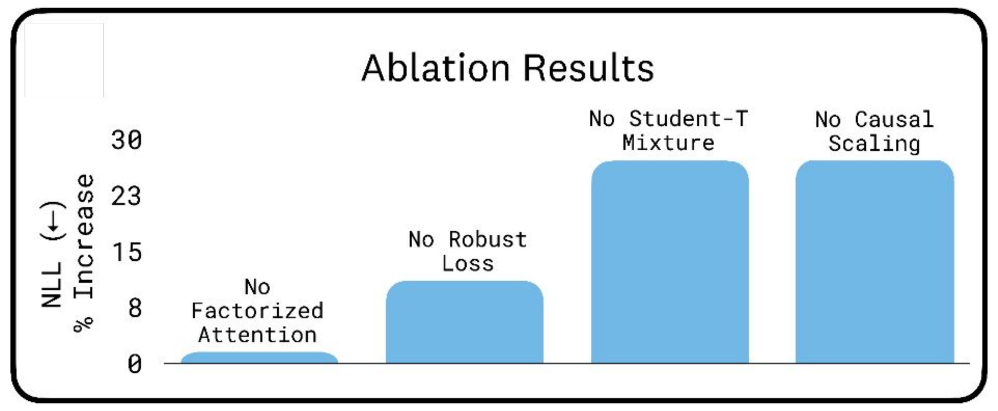
	Zero Shot					Full Shot				Baselines		
Metric	Toto	$\text{Moirai}_{\text{Large}}$	$\text{TimesFM}_{2.0}$	$\text{Chronos}_{\text{Bolt-Base}}$	TabPFN-TS	TEMPO	TTM-R2	PatchTST	TFT	Auto-ARIMA	Auto-ETS	Auto-Theta
MASE ↓	**0.673**	0.785	0.680	0.725	0.748	0.773	<u>0.679</u>	0.762	0.822	0.964	1.088	0.978
CRPS ↓	<u>0.437</u>	0.506	0.465	0.485	0.480	**0.434**	0.492	0.496	0.511	0.770	6.327	1.051
Rank ↓	**5.495**	10.330	8.412	<u>8.309</u>	8.402	8.897	10.103	10.268	11.629	21.608	25.134	24.134

BOOM

	Zero Shot							Baselines		
Metric	Toto	$\text{Moirai}_{\text{Base}}$	$\text{TimesFM}_{2.0}$	$\text{Chronos}_{\text{Bolt-Base}}$	Timer	$\text{Time-MoE}_{\text{Base}}$	VisionTS	Auto-ARIMA	Auto-ETS	Auto-Theta
MASE ↓	**0.617**	<u>0.710</u>	0.725	0.726	0.796	0.806	0.988	0.824	0.842	1.123
CRPS ↓	**0.375**	<u>0.428</u>	0.447	0.451	0.639	0.649	0.673	0.736	1.975	1.018
Rank ↓	**2.336**	<u>4.253</u>	5.155	5.447	9.370	9.381	10.317	9.16	10.956	11.712

Ablations

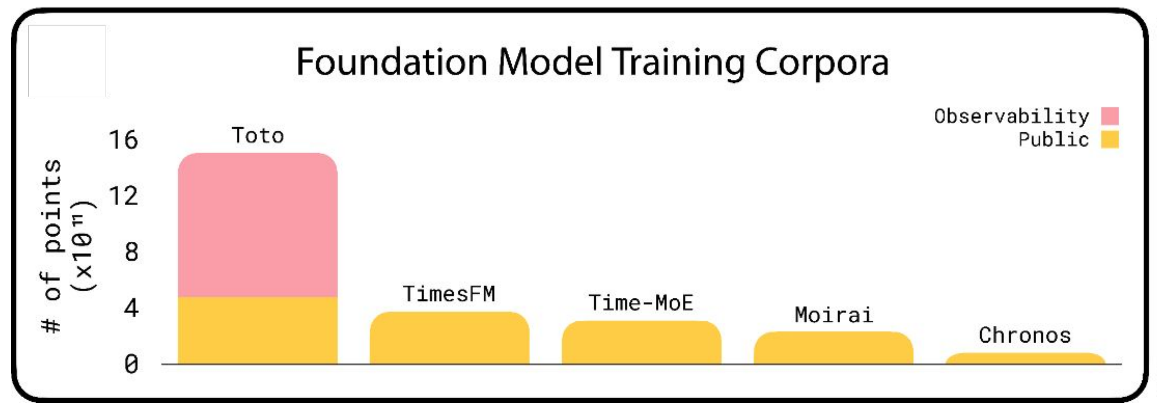
- Largest performance gains come from the Student-T mixture head and causal scaling (each causing ~27% NLL degradation when removed), with variate-wise attention and the Cauchy loss term providing smaller but still measurable improvements.



Pre-training Data

Total training: 2.36 trillion time points:

- 43% internal Datadog observability data
- 24% public (e.g., GIFT-Eval, Chronos)
- 33% synthetic time series
- Largest training corpus **(by 4–10×)**



Resources

Model Card

BOOM Dataset

Code

Come Join Us!

If this type of work interests you, consider applying to Datadog AI Research. We are growing our team of scientists and engineers in both New York City and Paris.