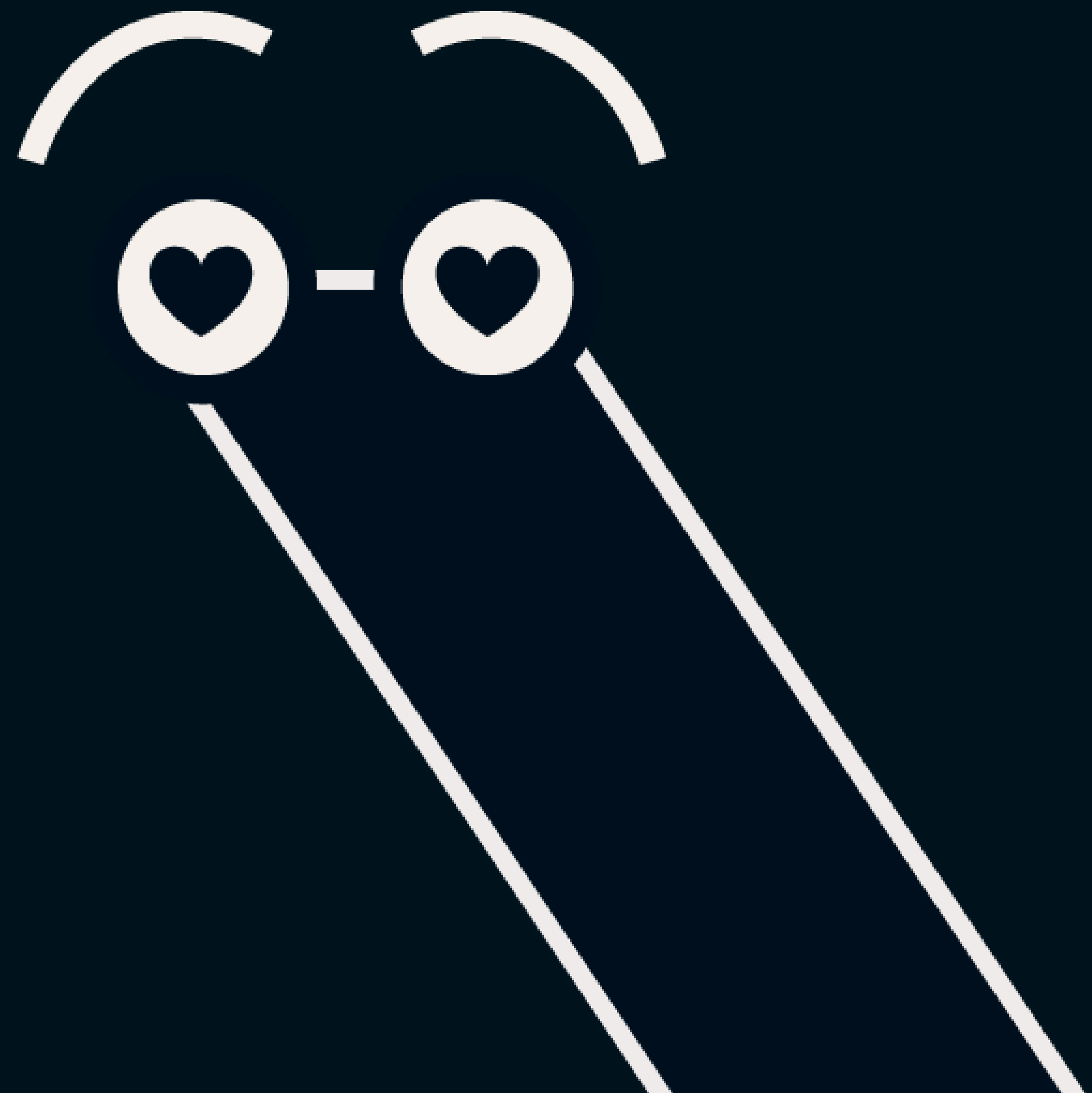
xtream

From Chickens to Transformers: The Evolution of Forecasting.



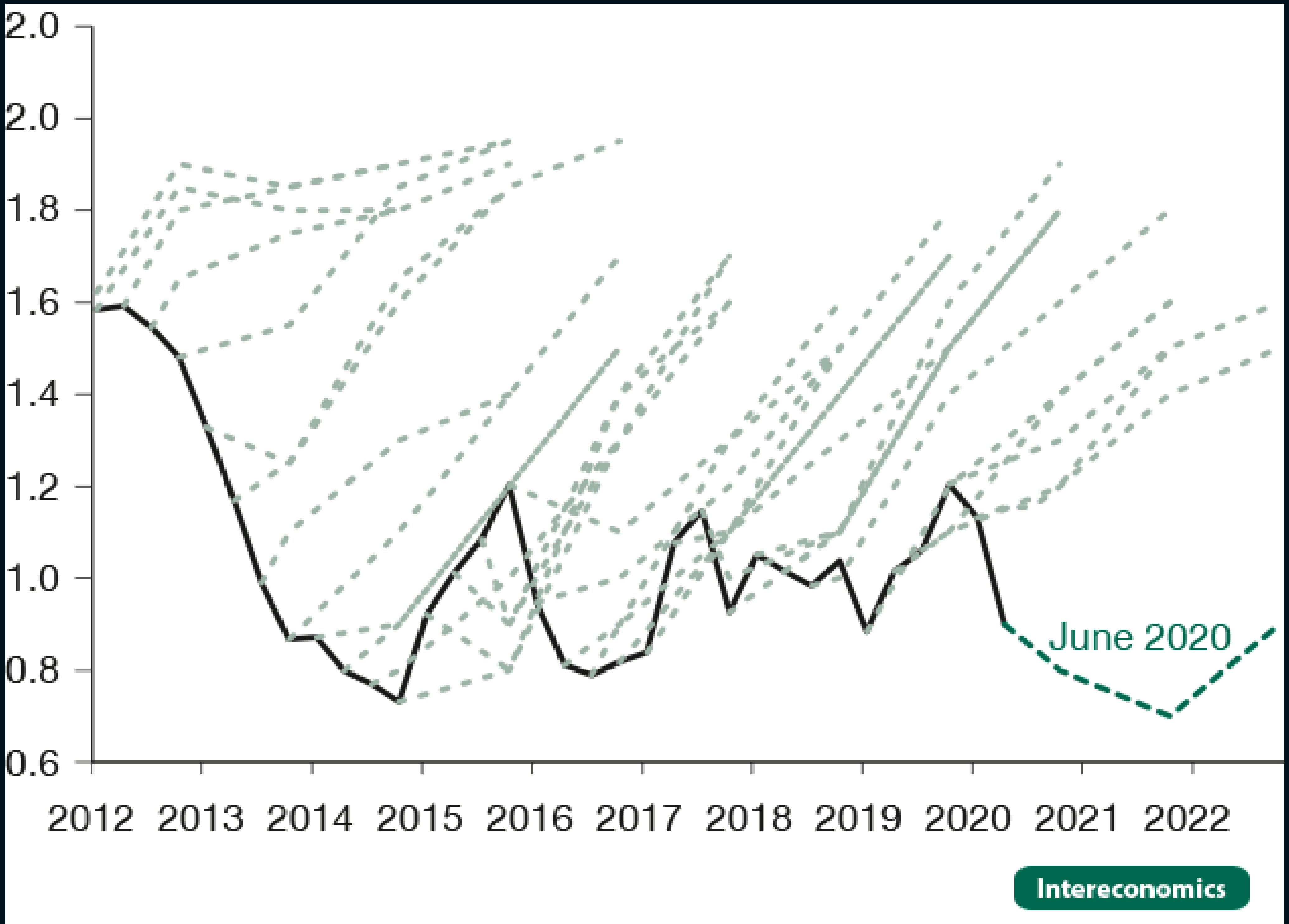
Emanuele Fabbiani

249 BC


HUYOT.

2000s

ECB Inflation Forecasts



"By 2005, it will become clear that the Internet's impact on the economy has been no greater than the fax machine's."

P. Krugman, Nobel Laurate


prime
amazon
prime
prime
prime

Up to 2015

Statistical Methods and ML Models (ARIMA, Random Forests)

https://www.amazon.science/latest-news/the-history-of-amazons-forecasting-algorithm

2015-2020

Deep Learning Models (RNN + CNN)

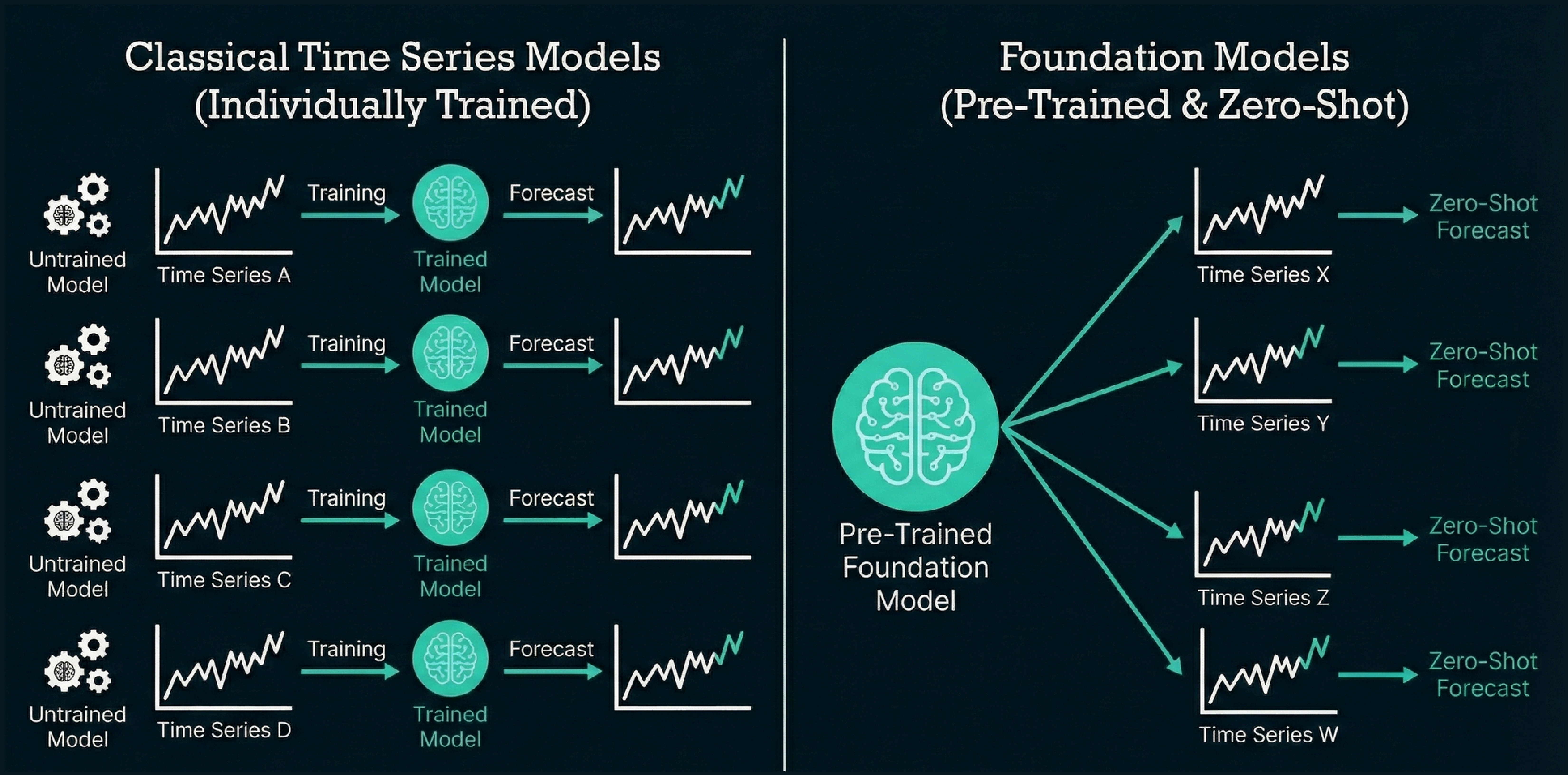
https://www.amazon.science/latest-news/the-history-of-amazons-forecasting-algorithm

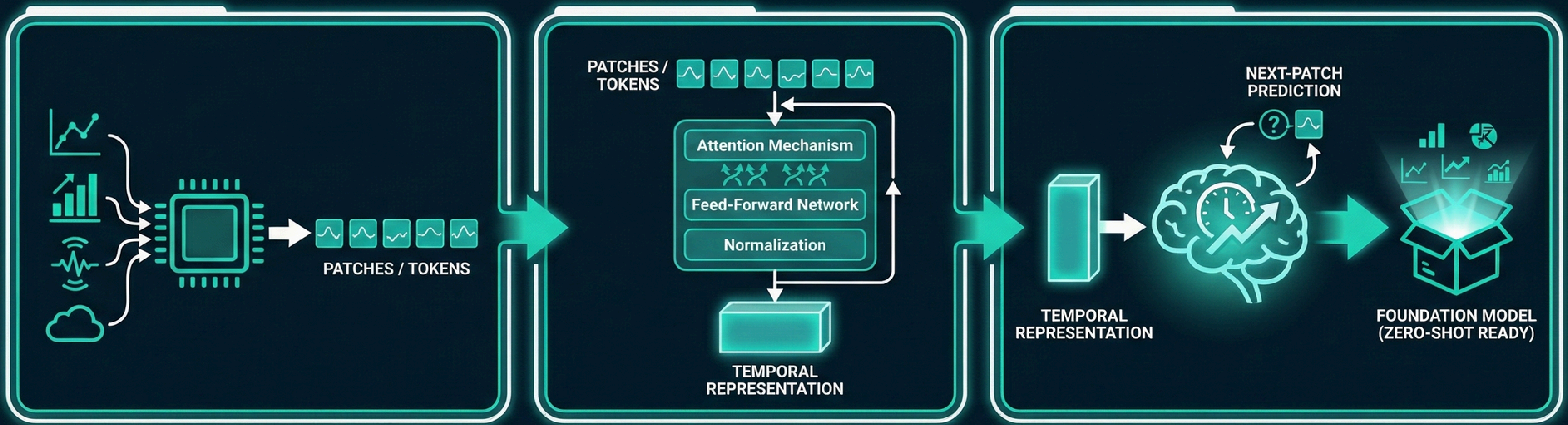
Since 2020

Transformers

https://www.amazon.science/latest-news/the-history-of-amazons-forecasting-algorithm

Foundation Models


Classical Time Series Models
(Individually Trained)
Untrained Model
Time Series A
Training
Trained Model
Forecast
Untrained Model
Time Series B
Training
Trained Model
Forecast
Untrained Model
Time Series C
Training
Trained Model
Forecast
Untrained Model
Time Series D
Training
Trained Model
Forecast
Foundation Models
(Pre-Trained & Zero-Shot)
Pre-Trained Foundation Model
Time Series X
Zero-Shot Forecast
Time Series Y
Zero-Shot Forecast
Time Series Z
Zero-Shot Forecast
Time Series W
Zero-Shot Forecast



1. Large-Scale Data Collection & Tokenization

Gather massive, diverse time series datasets. Break down sequences into uniform patches or tokens for model processing.

2. Transformer-Based Architecture Design

Construct a deep network (e.g., Transformer) to learn long-range dependencies and patterns within the data sequence.

3. Pre-training for General Patterns

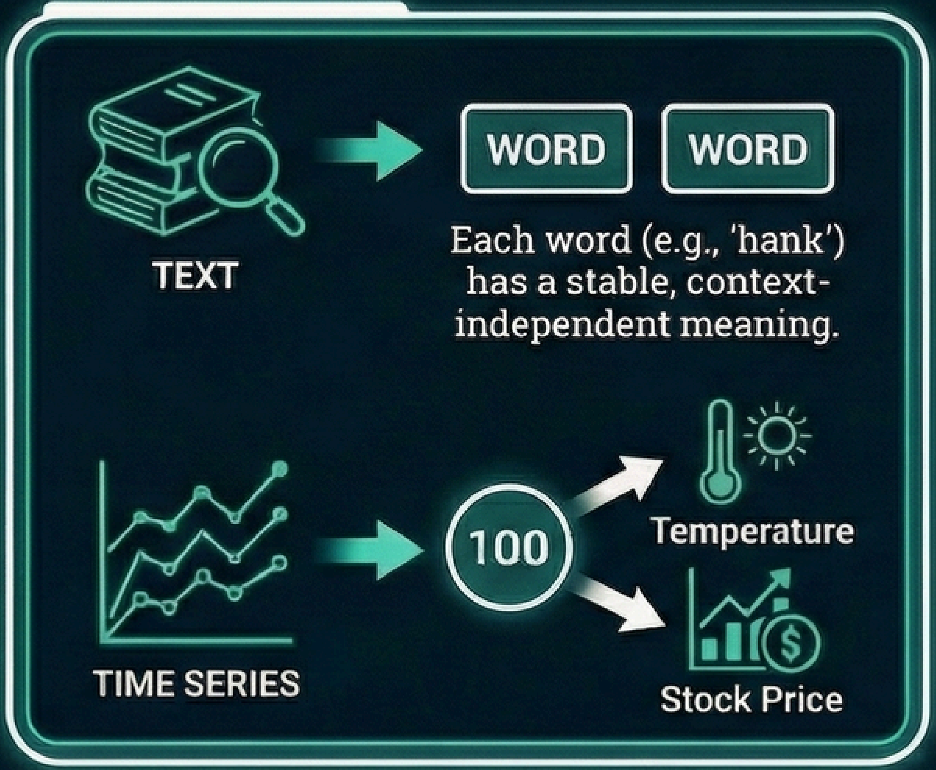
Train on the vast corpus using self-supervised tasks to learn universal temporal dynamics for immediate use or fine-tuning.

The Race to the Foundation Model

1. Google, TimesFM 2.5 (2025)
2. Amazon, Chronos 2 (2025)
3. IBM, StateFlow (2025)
4. Prior Labs, TabPFN-TS (2025)
5. Salesforce, Moirai (2024)

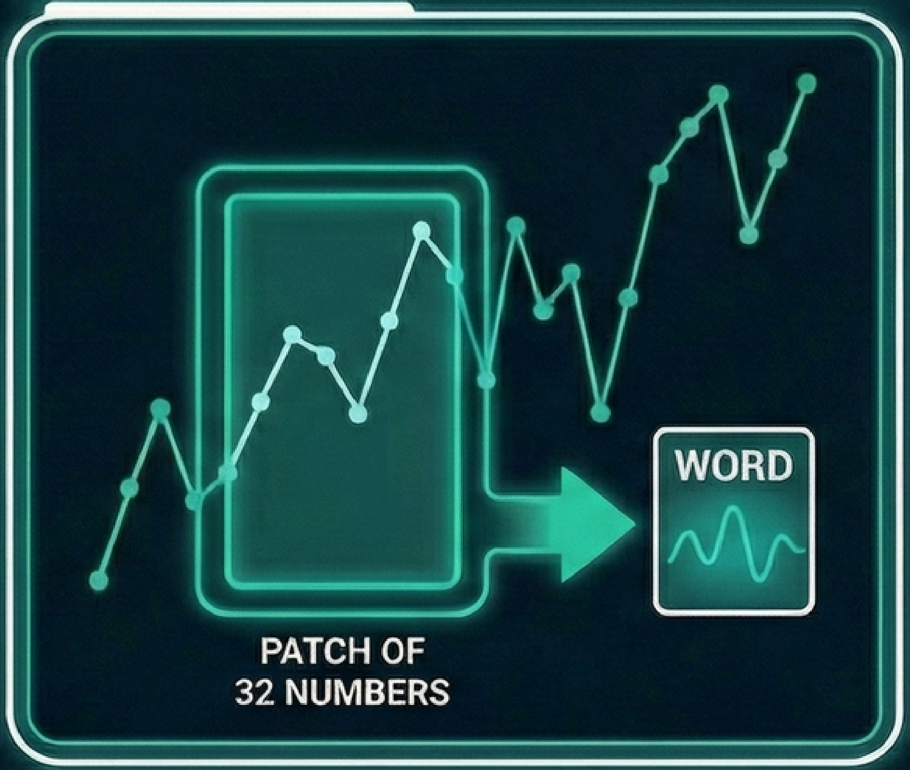
And Their Challenges

TWO POSSIBLE SOLUTIONS



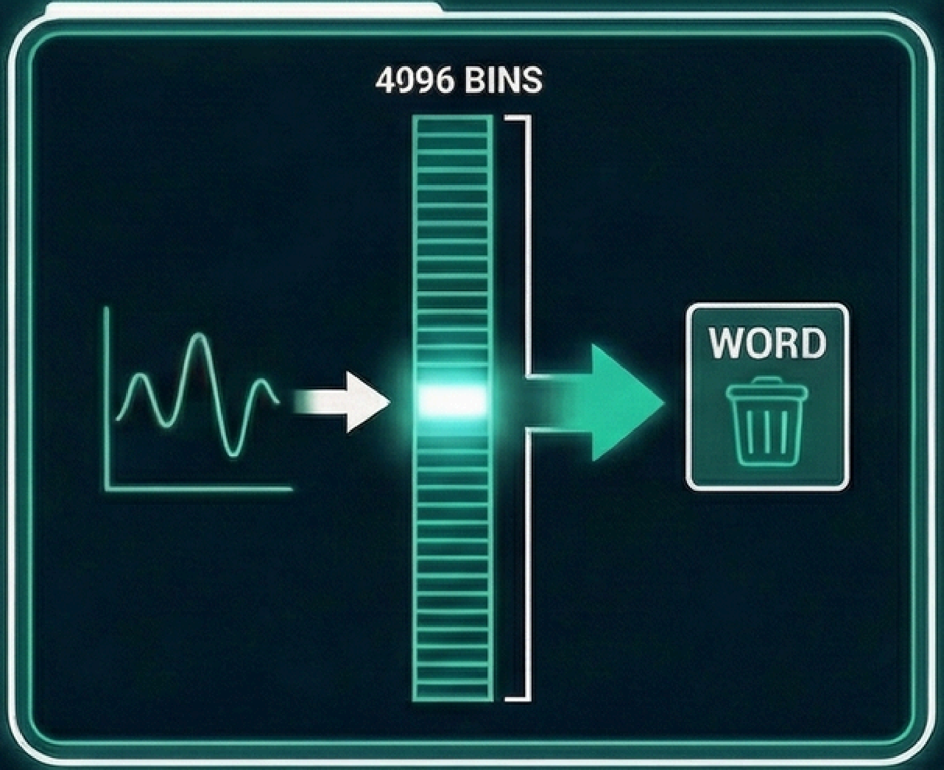
The 'Semantic' Challenge: Context-Dependent Meaning

The same number (e.g., '100') can have very different meanings depending on the specific context.



SOLUTION 1: Patching (Google, TimesFM)

Group consecutive points into a single patch (e.g., 32 numbers) to form a "word" capturing local context.



SOLUTION 2: Quantization (Amazon, Chronos)

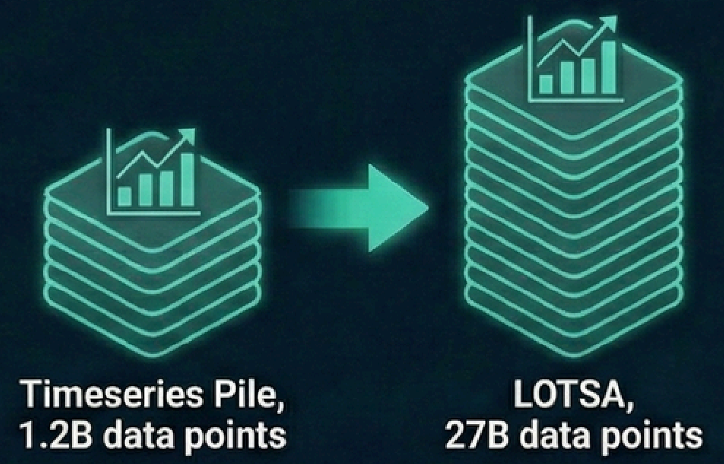
Quantize each time series into a fixed number of bins (e.g., 4096) and use each bin as a distinct "word".

TWO POSSIBLE SOLUTIONS

The 'Data Scarcity' Challenge: Obtaining Sufficient Data

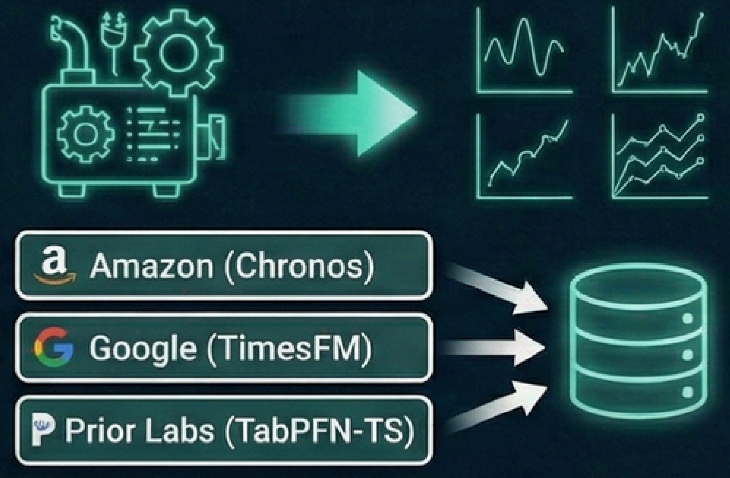
Obtaining sufficiently large and diverse time series datasets for robust training is difficult due to privacy, cost, and availability issues.

SOLUTION 1: Massive Data Collection



Aggressively collect vast quantities of diverse time series data from numerous sources. Examples include the Timeseries Pile (1.2 billion points) and the even larger LOTSA (27 billion points), providing extensive real-world examples.

SOLUTION 2: Synthetic Augmentation



Generate high-quality synthetic time series data to supplement real data, covering rare events and diverse scenarios. Proven effective by models from Amazon (Chronos), Google (TimesFM), and Prior Labs (TabPFN-TS).

Are Transformers Effective for Time Series Forecasting?

Ailing Zeng[1]*, Muxi Chen[1]*, Lei Zhang[2], Qiang Xu[1]
[1]The Chinese University of Hong Kong
[2]International Digital Economy Academy (IDEA)
{alzeng, mxchen21, qxu}@cse.cuhk.edu.hk
{leizhang}@idea.edu.cn

Abstract

Recently, there has been a surge of Transformer-based solutions for the long-term time series forecasting (LTSF) task. Despite the growing performance over the past few years, we question the validity of this line of research in this work. *Specifically, Transformers is arguably the most successful solution to extract the semantic correlations among the elements in a long sequence. However, in time series modeling, we are to extract the temporal relations in an* ordered set of continuous points. *While employing positional encoding and using tokens to embed sub-series in Transformers facilitate preserving some ordering information, the nature of the* permutation-invariant *self-attention mechanism inevitably results in temporal information loss.*

To validate our claim, we introduce a set of embarrassingly simple one-layer linear models named LTSF-Linear *for comparison. Experimental results on nine real-life datasets show that* LTSF-Linear *surprisingly outperforms existing sophisticated Transformer-based LTSF models in all cases, and often by a large margin. Moreover, we conduct comprehensive empirical studies to explore the impacts of various design elements of LTSF models on their temporal relation extraction capability. We hope this surprising finding opens up new research directions for the LTSF task. We also advocate revisiting the validity of Transformer-based solutions for other time series analysis tasks (e.g., anomaly detection) in the future. Code is available at:* `https://github.com/cure-lab/LTSF-Linear`.

ergy management, and financial inve
several decades, TSF solutions have
sion from traditional statistical metho
and machine learning techniques (
deep learning-based solutions, e.g.,
works [15] and Temporal Convolutio

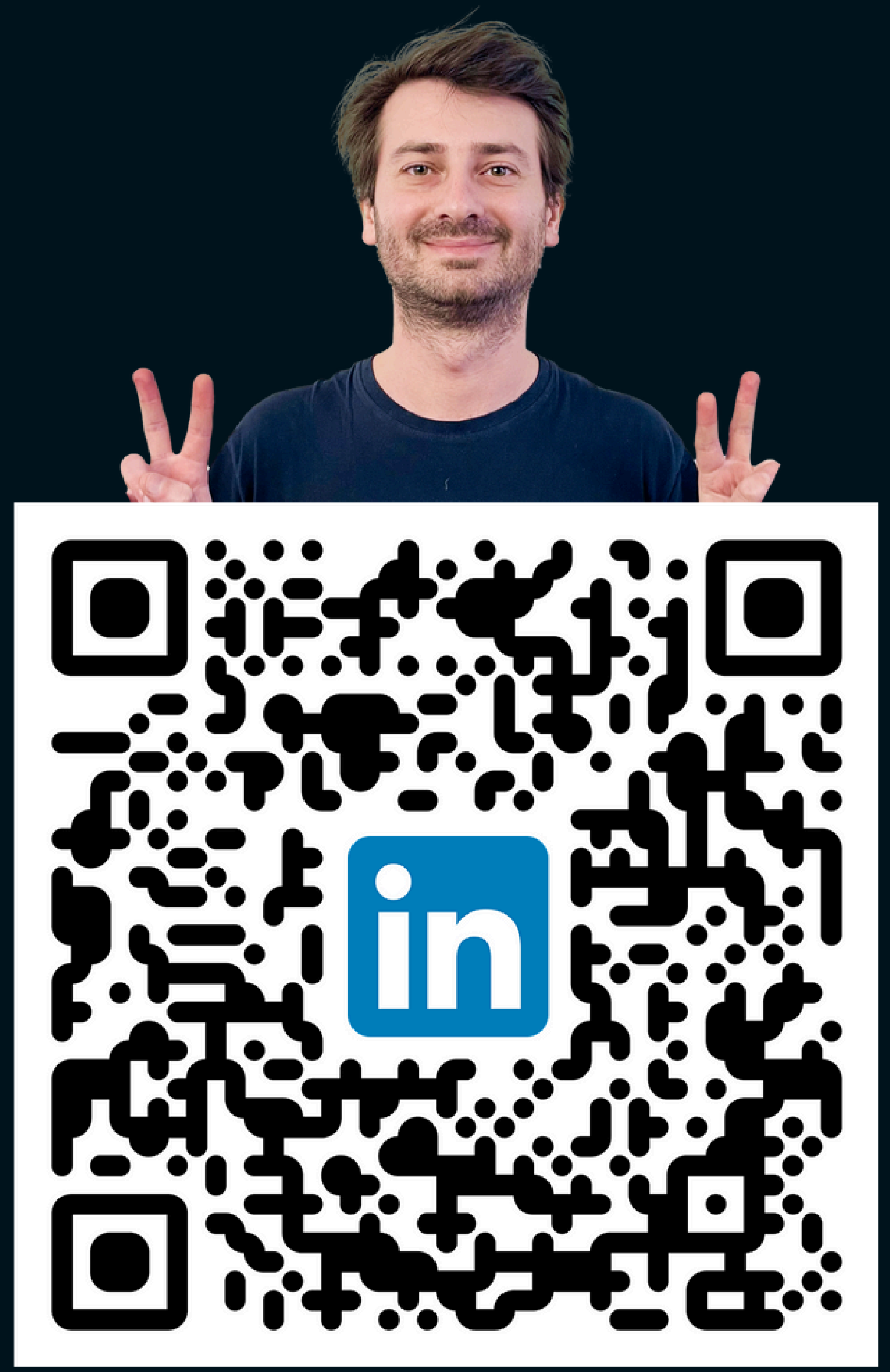
Transformer [26] is arguably the
quence modeling architecture, demo
performances in various applications
guage processing (NLP) [7], speech
computer vision [19, 29]. Recently,
surge of Transformer-based solutions
ysis, as surveyed in [27]. Most n
focus on the less explored and challe
series forecasting (LTSF) problem, i
(NeurIPS 2019), Informer [30] (AA
Autoformer [28] (NeurIPS 2021), Py
2022 Oral), Triformer [5] (IJCAI 202
former [31] (ICML 2022).

The main working power of Tra
multi-head self-attention mechanism
able capability of extracting semanti
elements in a long sequence (e.g.,
patches in images). However, self-att
invariant and "anti-order" to some ex
ious types of positional encoding tec
some ordering information, it is still i
poral information loss after applying
of them. This is usually not a serious
rich applications such as NLP, e.g.,
of a sentence is largely preserved ev

Ego Slide

Emanuele Fabbiani

Head of AI @ xtream



Co-founder of xtream and Reaidy (both **acquired** by TeamSystem)

Professor in AI at Catholic University of Milan

PhD in Applied AI

Speaker at AMLD Lausanne, ODSC London, PyData Berlin, Milan, Paris, WeAreDevelopers Berlin, PyCon Italy, Swiss Python Summit.