

Are we ready for Time Series Foundation Models in economics?

Lessons from an ESCB/SSM Hackathon

Ada Lovelace Algorithmic Lectures | DATA LAB

Nabil Bouamara
Data Science Center

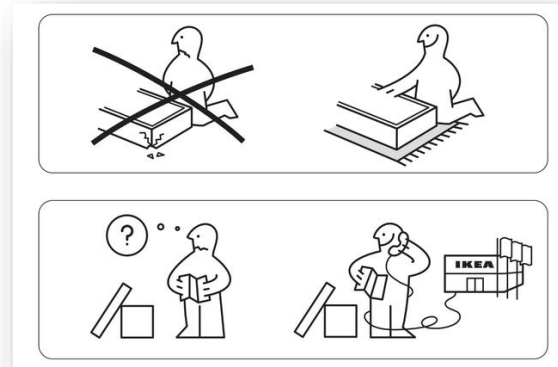
March 26, 2026

The burning question

Can economic forecasting move:

from
***carefully specified, task-specific
econometric models***

to
pretrained, reusable forecasting systems?



Why this matters in central banking

Forecasting is recurrent

Forecasting is operational

Forecasting is high-stakes

Accuracy matters, but so do:

*robustness, interpretability,
reproducibility, behaviour under
revisions and breaks*

Many workflows are still task by task

ECB holds interest rates at 2% as energy prices soar

Polymakers warn of 'material' impact on near-term inflation



Classic forecasting logic

Classical central-bank forecasting

start from a specific variable

define the forecasting question

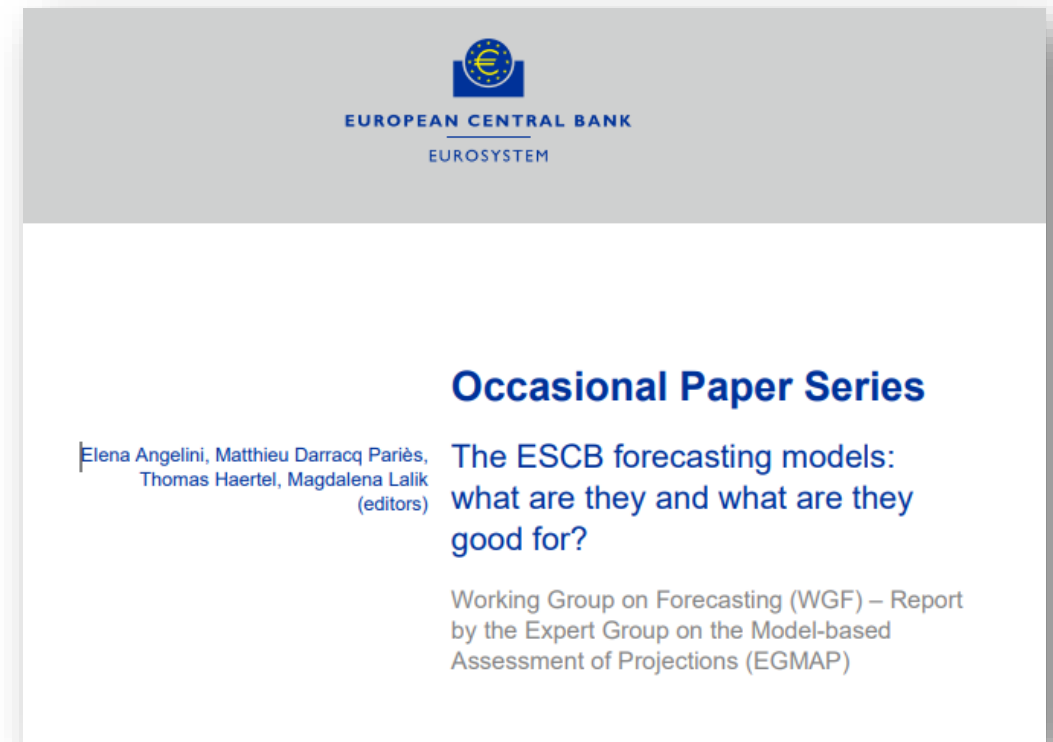
choose a suitable model

bring in domain knowledge early

estimate parameters on specific dataset

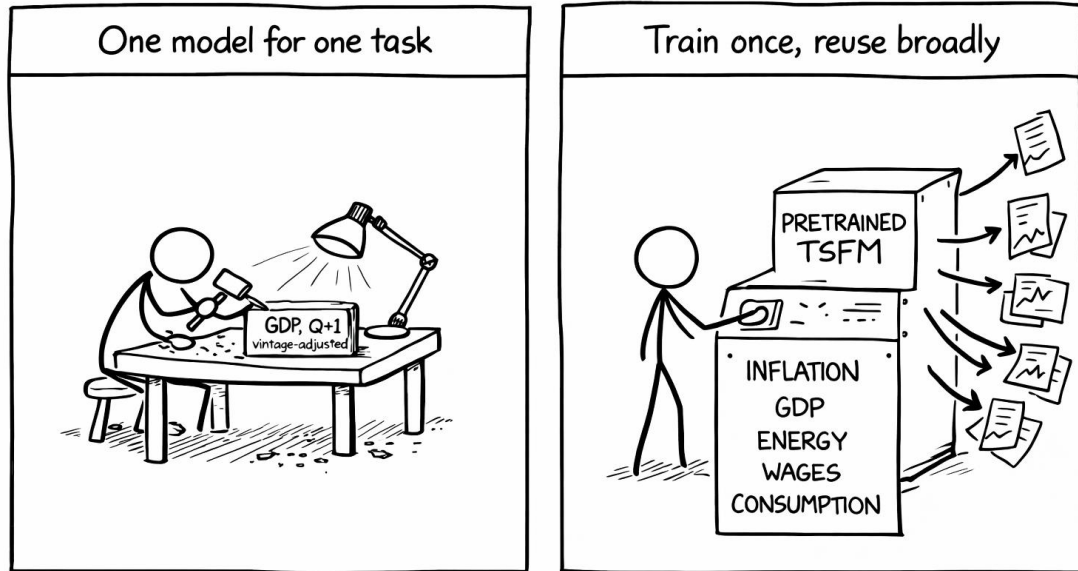
forecast using the estimated model

*Examples: ARIMA / SARIMA, factor models,
bridge equations, state-space models,
Bayesian VARs, DSGE, ...*



*Forecasting in central banking is already
a mature model ecosystem*

Foundation model forecasting logic



Foundation-model forecasting

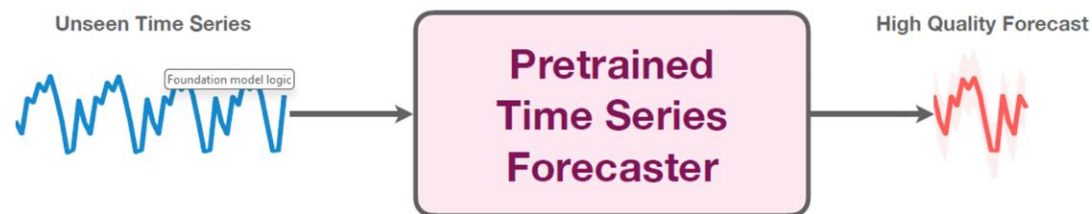
*pretrain on many time series
learn reusable temporal patterns across
diverse time series
reuse on new datasets
apply off-the-shelf to your specific
forecasting task, no estimation needed
bring in task-specific knowledge later*

*Examples: Chronos / Chronos-2
(Amazon), TimesFM (Google), TimeGPT
(Nixtla), Lag-Llama, Moirai (Salesforce)*

The promise of TSFMs, according to Amazon

Time Series Foundation Models

Time Series tasks have diverse **patterns, frequencies, history lengths, prediction horizons, missing values, ...** → *long trial-and-error development cycles*



*... we could develop a **pretrained forecaster** that performs well on unseen time series tasks?*

aws

Let's put those side by side

Econometric Models

- Task-specific
- Requires domain expertise
- Limited by data availability
- Need re-estimation for new data
- Interpretable parameters

Foundation Models

- General-purpose
- Learn from massive datasets
- Transfer learning across domains
- Zero-shot forecasting
- more “black-box” like

How does the model “see” the series?

*Models do not see raw reality;
directly.*

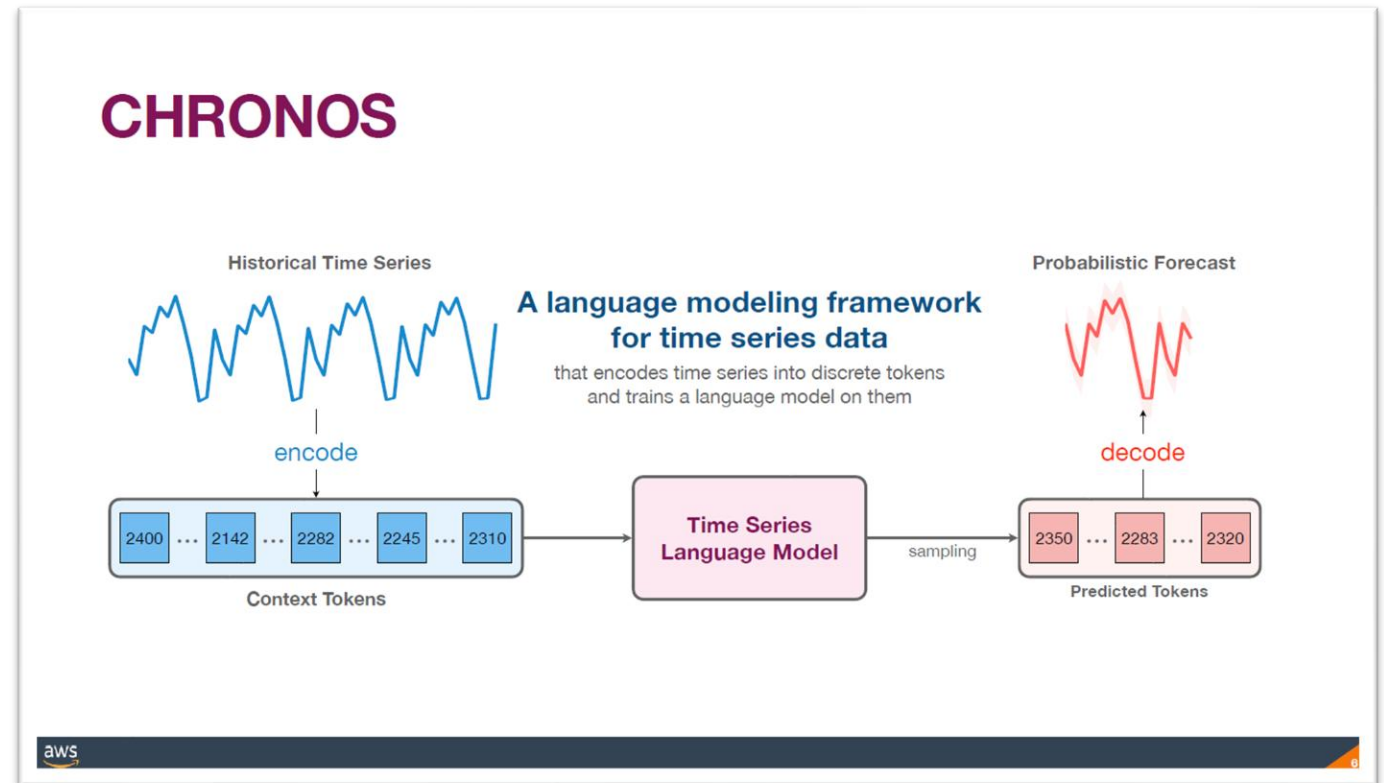
*They see **a representation** of it*

***Possible time-series
representations:***

tokens

patches

continuous embeddings

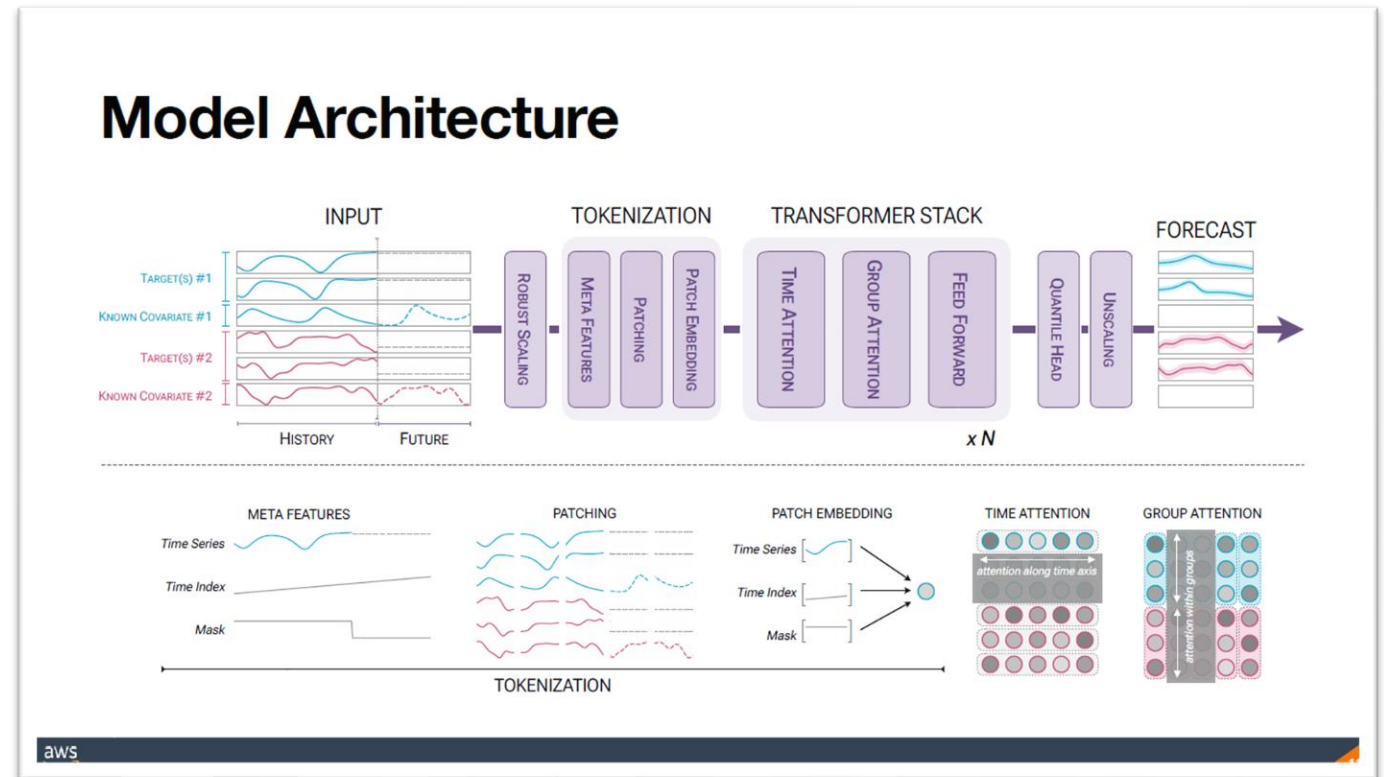


How does the model “see” the series?

*Models do not see raw reality;
directly.*

*They see **a representation** of it*

*The tokenization inside the full
architecture of Chronos 2.0*



How does the model “see” the series?

*Models do not see raw reality;
directly.*

*They see **a representation** of it*

The promise

Introducing Chronos-2

Chronos 2.0

A **unified** model supporting...

- **cross learning**
infers better using multiple items in the batch
- **multivariate forecasting**
infers dependencies across different related series
- **exogenous variables**
infers how target time series depend on external factors

...via in-context learning.

aws

How does the model “see” the series?

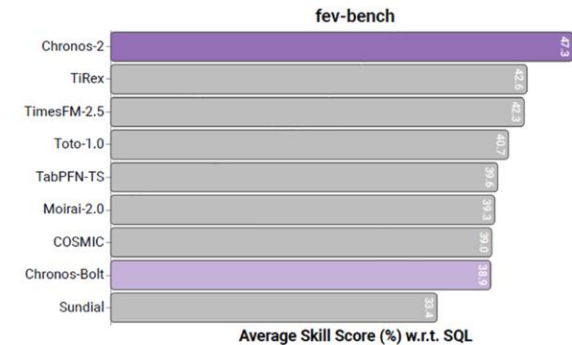
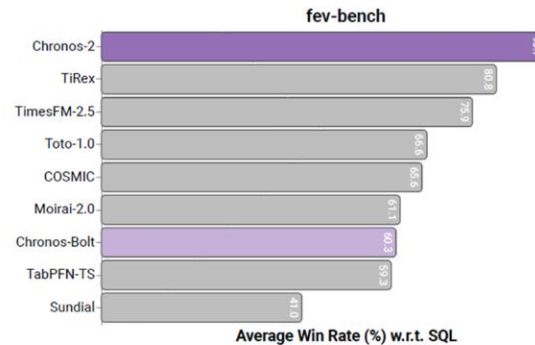
*Models do not see raw reality;
directly.*

*They see **a representation** of it*

Compared to others...

fev-bench Results

fev-bench: Comprehensive benchmark with 100 tasks including univariate, multivariate and covariate-informed forecasting



Results with respect to the scaled quantile loss (SQL) which evaluates probabilistic forecasting performance

aws

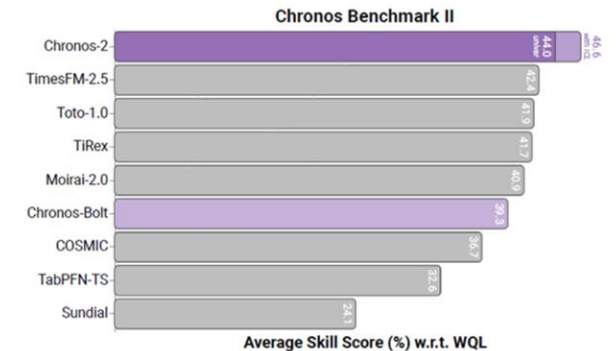
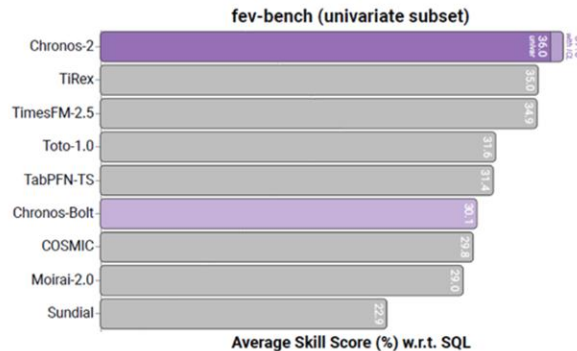
How does the model “see” the series?

*Models do not see raw reality;
directly.*

*They see **a representation** of it*

Compared to others...

ICL on Univariate Tasks



How does the model “see” the series?

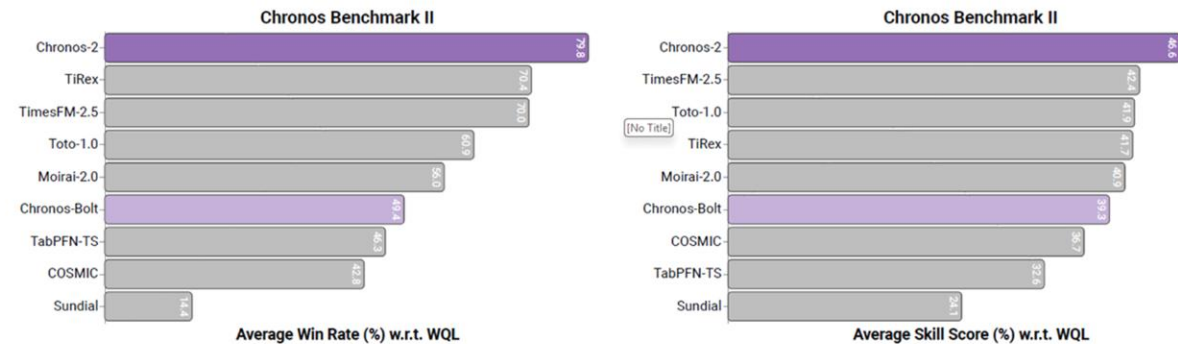
*Models do not see raw reality;
directly.*

*They see **a representation** of it*

Compared to others...

Chronos Bench II Results

Chronos Bench II: Benchmark with 27 tasks, with the majority being short-context datasets



Results with respect to the weighted quantile loss (WQL) which evaluates probabilistic forecasting performance

aws

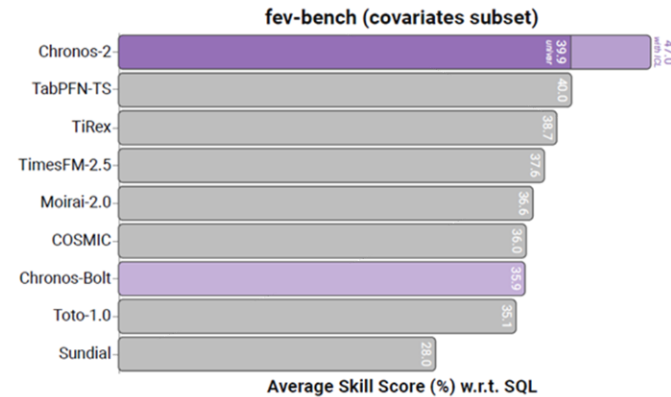
How does the model “see” the series?

*Models do not see raw reality;
directly.*

*They see **a representation** of it*

Compared to others...

ICL on Covariate-informed Tasks



Why time series are not just another language

Language

natural tokens

syntax

semantics

Time series

no natural vocabulary

meaning depends on context

the same shape can reflect different mechanisms

So why are people excited anyway?

Potential gains:

faster, stronger baselines

less per-task engineering

better transfer in low-data settings

easier reuse across domains

Old default: historical mean, random walk, AR(1), ...

The hackathon challenge

5th ESCB/SSM hackathon — winning team, 2025

*24 hours, 64 participants, 27
institutions*

ESCB motivation

*Modernize the analytical toolkit in
light of rapid developments in AI*

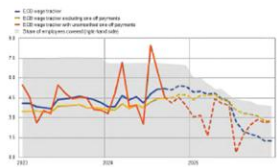


The hackathon challenge

What the challenge combined
news-based signals
traditional indicators
foundation models
reusable deployment tools

Challenge overview – Objectives

Can news help forecasting wage pressures and investment decisions?



Improve the short-term forecasting of wage pressures and corporate investment.

GDELT high-frequency data and traditional indicators



Learn how real-time media signals can anticipate labour market tensions, wage pressures and firm investment decisions.

AI Foundation models



Develop models based on selected AI pre-trained models with optimised performance for the specific dataset.

Development platform in ECB cloud



Build a safe, state-of-the art, cost-conscious platform, which can anticipate future needs for predictions across diverse business areas.

Reusability for diverse use cases

10

The hackathon challenge

Traditional economic input

Challenge overview – Economic data (input)

60

Economic variables

- Publicly available (ECB, SDW, Eurostat, EC)
- Focus on one of the 4 largest euro area countries (Germany, France, Italy, Spain)
- Quarterly and monthly frequencies

Wages

- Wage and compensation indicators
- Price and Inflation indicators
- Economic sentiment and expectations
- Labour market indicators
- Economic indicators

Investment

Quarterly investment in machinery and equipment

Monthly indicators:

- Production
- Sales and Trade
- Business confidence/sentiment indicators, expectations

11

The hackathon challenge

Why news data mattered

GDELT: turning news into signals

Challenge overview – media signals data (input)

What is the GDELT project?

- High-frequency, media-derived indicators from the Global Database of Events, Language and Tone
- Project running since 2013
- From publicly available web sites **without a fee**
- 16TB of news data (adding around 2TB/year)
- Real-time signals from GDELT can help in nowcasting macroeconomic indicators (e.g. unemployment rate).

You will be given example code to create aggregate series (using distributed computing) but are expected to develop your own approach!

Dataset

- Reduced version (4TB) of the GDELT Global Knowledge Graph (1 record = 1 article)
- only articles mentioning at least one of the four largest Euro area countries
 - only a subset of columns (see next slide)

What should you do with that?

- **Create aggregate measures** at quarterly/monthly frequency.
 - Media attention: number of articles, total word count, count of occurrences of specific themes
 - Media sentiment: average sentiment/positivity/negativity
- **IMPORTANT:** control for overall amount of news

14

The hackathon challenge

Translating news into economically meaningful features

Challenge overview – media signals data (input)

Example: Sky news article on EU GDP release

- **Date:** 2018-02-14 16:15:00 (not exactly the publication time)
- **Themes:** ['USPEC_POLICY1', 'EPU_ECONOMY', 'EPU_ECONOMY_HISTORIC', 'ECON_STOCKMARKET', 'ECON_INFLATION', 'WB_1104_MACROECONOMIC_VULNERABILITY_A', 'ND_DEBT', 'WB_442_INFLATION']
- **Locations:** [(1, Germany, GM, GM, 51.5, 10.5, GM), (1, Bulgaria, BU, BU, 43, 25, BU), (1, United Kingdom, UK, UK, 54, -4, UK), (1, France, FR, FR, 46, 2, FR), (1, Spain, SP, SP, 40, -4, SP)] - note different country codes according to FIPS standard
- **Countrycodes:** ['DE', 'FR', 'ES'] in ISO-2-C
- **Tone:** [1.8, 2.7, 0.9, 3.6, 15.8, 2.3, 208] (tone, positive score, negative score, polarity, active reference density, self/group reference density, wordcount)
- **Persons:** None
- **Organisations:** [European union]
- **Amounts:** [(28, 'countries put', 162), (19, 'countries using the single', 470)]
- **GCAM:** excerpt: [.. 'c41.1':23], ['c6.6':1]...]
(c41.1: wordcount "Central Bank Financial Stability - positive", c6.6: wordcount "Loughran and McDonald Financial Sentiment - Uncertainty")
- **Sourcecommonname:** "sky.com"

European Union economy records fastest economic growth in a decade

The EU economy grew at its fastest rate in 10 years in 2017, registering a 2.5% increase on the year before. It is the highest annual growth since 2007, when the economy expanded by 2.7%.

The bloc of 28 countries put in a strong performance in the final quarter of the year, growing 0.6%, mainly driven by good economic results from Germany, Spain and France. However, Eastern European states including Slovakia and Bulgaria also contributed.

The figures, released by the European Union statistics office Eurostat, showed that the eurozone - which encompasses the 19 countries using the single currency - also had a good year, with GDP growth of 2.5%. The strength of the figures show the extent to which the UK has been trailing its European counterparts.

UK GDP grew by 1.8% overall in 2017, the slowest rate of increase since 2012.

The EU figures kept the FTSE 100 and European markets in a good mood for most of the morning, with each of the main indices up by around 0.6%. That progress reversed however after US inflation increased by more than expected, raising fears of that a rate rise sooner than investors would like.

15

The hackathon challenge

*Some promises were made
(generalization, efficiency, and
productivity); but how much of it
survives in practice?*

Challenge overview – TSFMs (input)

Time Series Foundation Models (TSFMs) represent a recent and fast-developing paradigm that brings the foundation model philosophy (as seen in large language models) to the time series domain. These models are pre-trained on vast collections of heterogeneous time series data, often across sectors, frequencies, and sources, and then adapted (fine-tuned or prompted) for specific forecasting, classification, or anomaly detection tasks.

Key characteristics and benefits

- **Large-scale pretraining:** trained on massive, heterogeneous time series datasets.
- +
- **Universal architectures:** Transformer or state-space hybrids, capturing complex temporal dependencies.
- +
- **Representation learning:** extract latent embeddings encoding trends, seasonality, volatility, non-linearities and regime shifts. Transfer knowledge across domains, countries, and frequencies

GENERALIZATION

- **Adaptability:** few-shot or zero-shot generalization via prompts or light fine-tuning.
- +
- **Reduced feature engineering:** minimal preprocessing or manual model specification

EFFICIENCY

- **Automation & Scalability:** handle thousands of time series with text, tabular, or other modalities.
- +
- **Robustness:** handle missing, noisy, or irregularly sampled data.

DATA PRODUCTIVITY

Times series foundation models for the hackathon

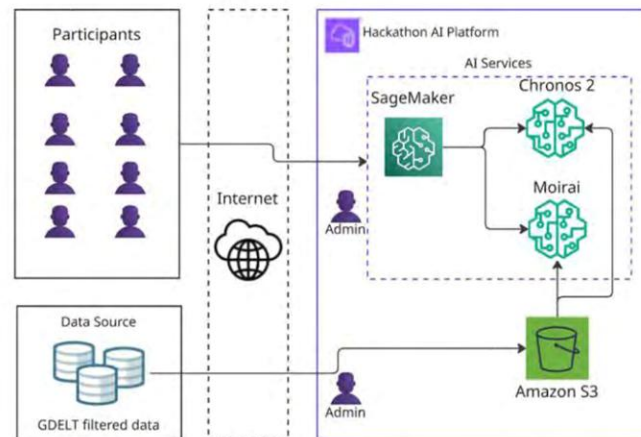
- Evaluation of top tier foundation models (approx. 10 models)
- Collaboration with Amazon Chronos research and development teams
- Selection of 2 models (Moirai 2.0), (Salesforce), Chronos 2.0 - multivariate (Amazon), live since October 2025)
- Caution in using only one model, security, sovereignty.
- Selection of only public data series
- **WARNING:** Interpretability

16

The hackathon challenge

It's not only about models; it's also about infrastructure.

AWS / Knowledge base & development platform

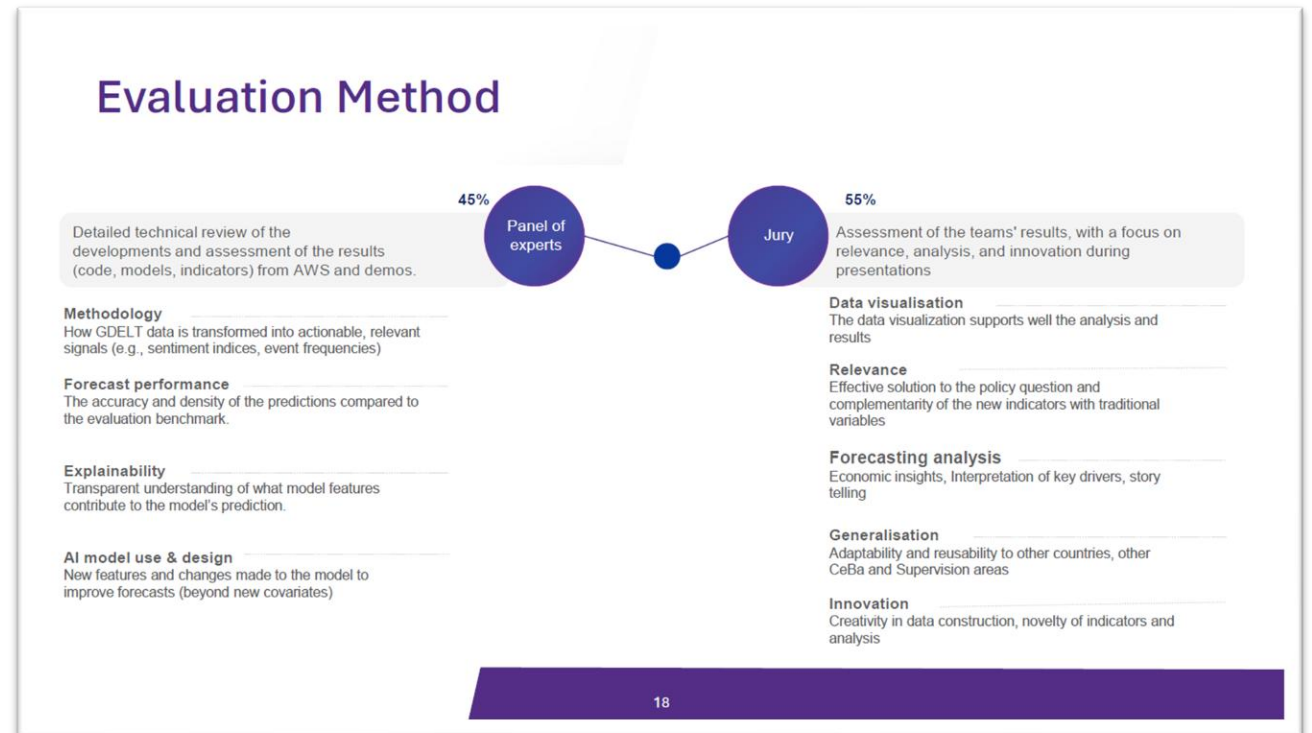


- Secure access to the accounts
- Resource separation between teams
- 2 AI TSFM available

20

The hackathon challenge

The hackathon was not judged on model accuracy alone.



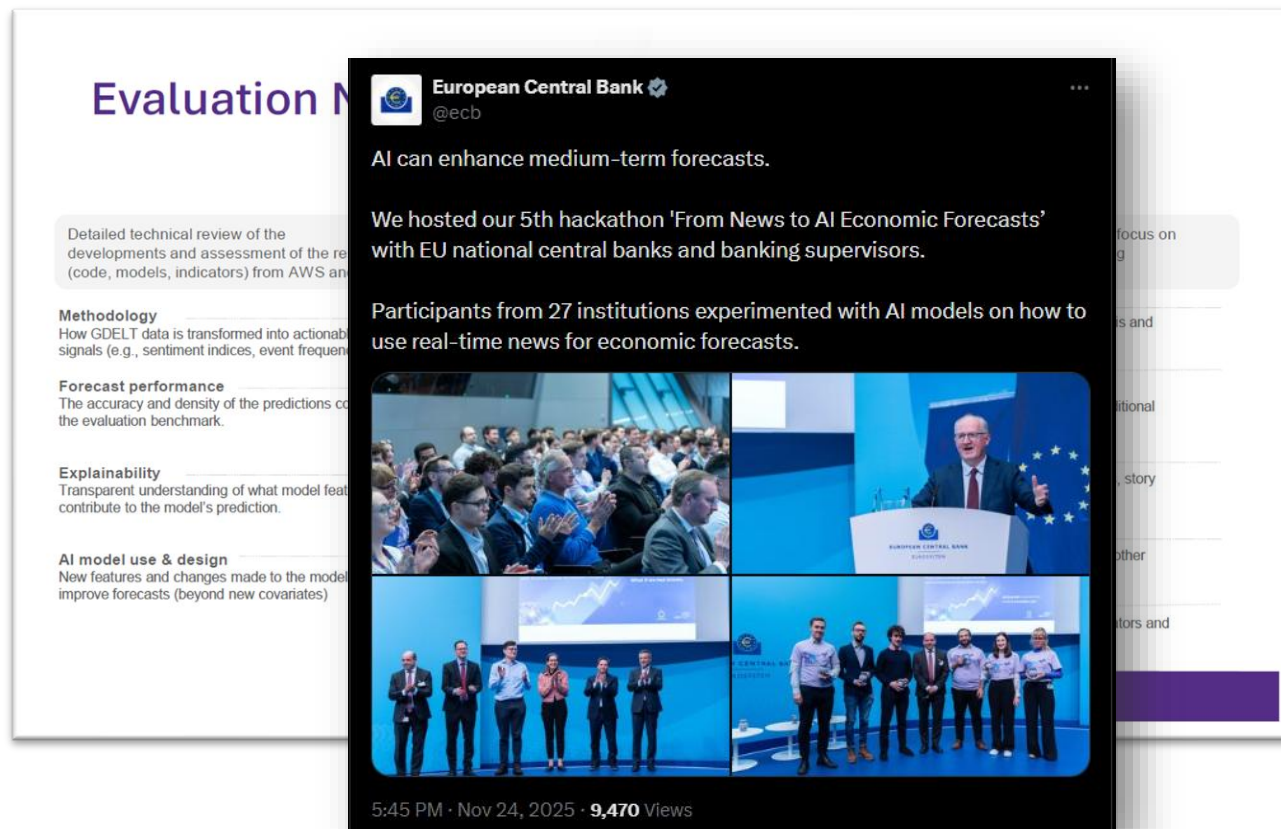
The hackathon challenge

Winning idea: make the black box more interpretable

Question: What is the real-time value-added of news?

Method: compare forecasts with and without news, using 'countertextuals'

Interpretation: "What would the model have predicted if it had already known this news?"



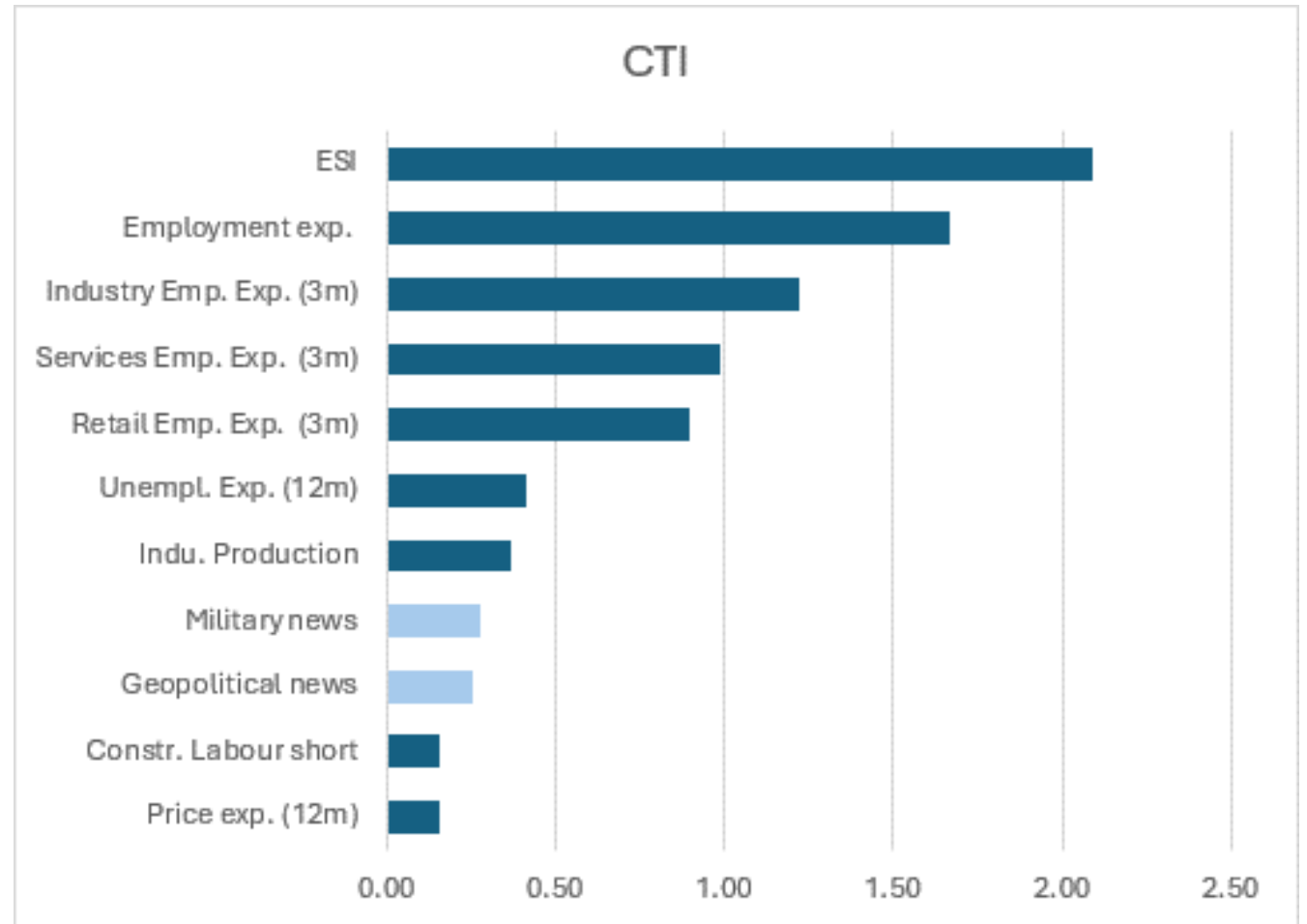
The hackathon challenge

Winning idea: make the black box more interpretable

Question: What is the real-time value-added of news?

Method: compare forecasts with and without news, using 'countertextuals'

Interpretation: "What would the model have predicted if it had already known this news?"



The hackathon challenge

Winning idea: make the black box more interpretable

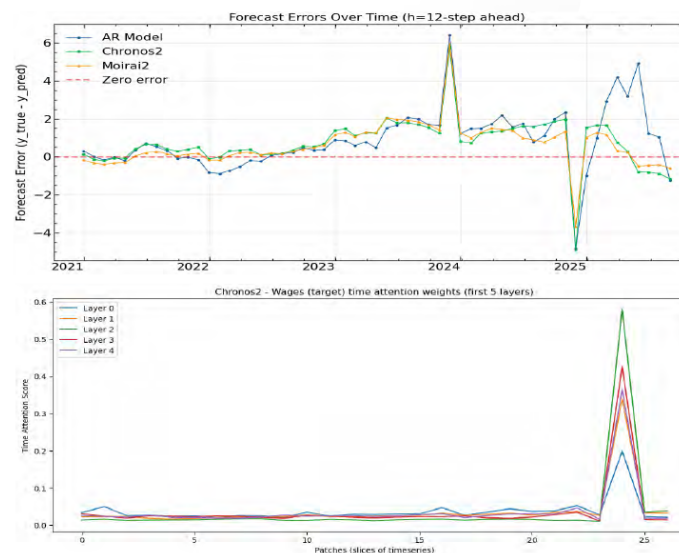
Question: What is the real-time value-added of news?

Method: compare forecasts with and without news, using 'countertextuals'

Interpretation: "What would the model have predicted if it had already known this news?"

What did others do?

Time Attention Layer



- There are 2 spikes in forecast errors in the forecast windows, a **data error** (2024) and a **model error** (2025)
- The Chronos 2 time attention weights, indicators of “autocorrelation” contribution to the target series
- The spike in probability across the layers at patch 24 suggests that the **model is picking up its own forecasted error** in its own time series

The hackathon challenge

Winning idea: make the black box more interpretable

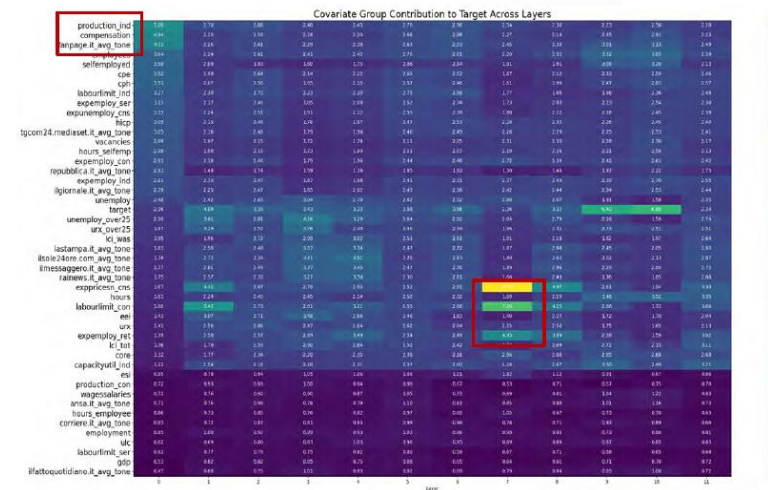
Question: What is the real-time value-added of news?

Method: compare forecasts with and without news, using 'countertextuals'

Interpretation: "What would the model have predicted if it had already known this news?"

What did others do?

Group Attention Layer



*Multiplied by 100 for visibility but the numbers represent probabilities and sum to 100 at the column level.

- At lower layers, we observe relatively high weights attributed to conventional economic indicators such as total **production in industry**, total **compensation of employees** followed by the **average tone of fanpage.it**, an investigative journalism site
- At higher layers, we observe expected prices to consumers, labour shortages in construction, employment expectations in the retail sector

The hackathon challenge

Winning idea: make the black box more interpretable

Question: What is the real-time value-added of news?

Method: compare forecasts with and without news, using 'countertextuals'

Interpretation: "What would the model have predicted if it had already known this news?"

What did others do?

LoRA concept

Training & Evaluation Pipeline

- Trains with AdamW on LoRA parameters only (base model remains frozen).
- Logs quantile pinball loss for train/validation at every epoch.
- Performs a final forward-only evaluation on the test set.
- Saves all metadata (loss curves, split sizes, hyper-parameters, test path) alongside the adapter checkpoint for reproducibility and comparison notebooks.

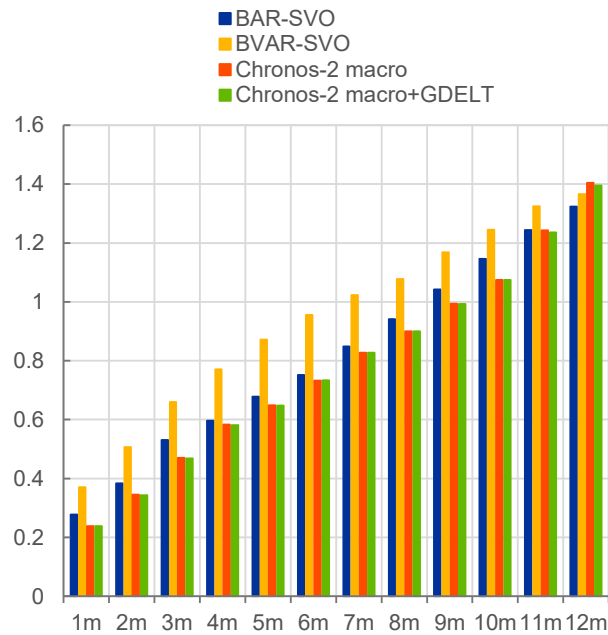
Hyper-parameters (CLI-Configurable)

- Context length: 40, prediction horizon: 4, stride: 1.
- Batch size: 16, epochs: 3, LR: 2e-4, weight decay: 0.
- Validation/test ratios: 0.1 / 0.1.
- LoRA config: rank = 8, alpha = 16, dropout = 0.05.
- Default LoRA targets: q_proj, k_proj, v_proj, out_proj.
- All parameters (dataset path, model, device, seed, output directory, series filters, etc.) are fully overrideable via CLI flags.

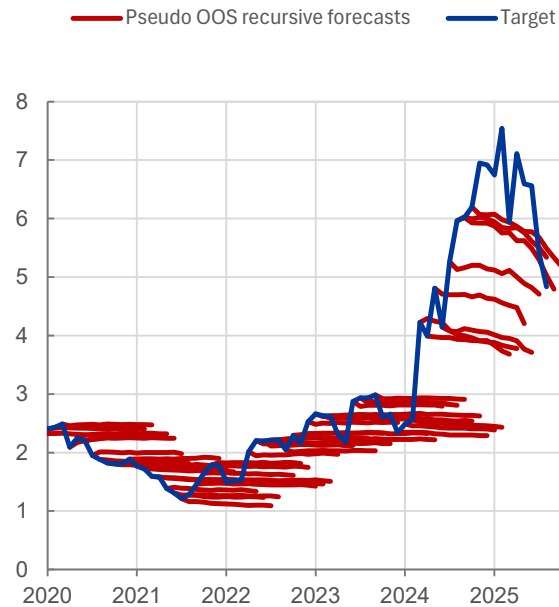
What still seems unresolved

useful enough to matter

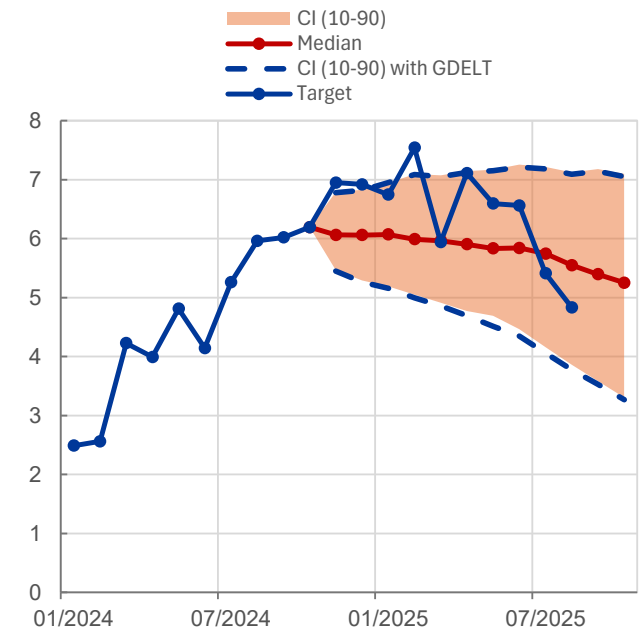
*not yet understood well enough to relax our standards
about interpretability and policy use*



Mean Absolute Forecast Error



Prediction



Balance of Risk around Projection