DIGITAL FINANCE

This project has received funding from the Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101119635

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

State Secretariat for Education, Research and Innovation SERI

Funded by the European Union

MARIE CURIE ACTIONS

DIGITAL

Advances in XAI

DIGITAL

Funded by
the European Union

Reading

Review of Advencements

- Reil, J. et al. (2025). Explaining the Unexplained: A Systematic Literature Review on Explainable Artificial Intelligence (XAI) in (Financial) Time Series Forecasting
- Altukhi, Z. M., Pradhan, S., & Aljohani, N. (2025).
  A systematic literature review of the latest advancements in explainable artificial intelligence. *Technologies, 13*(3), 93. https://doi.org/10.3390/technologies13030093

Model Monitoring & Deployment XAI

- Koebler, A., Decker, T., Lebacher, M., Thon, I., Tresp, V., & Buettner, F. (2023). Towards explanatory model monitoring. Proceedings of the NeurIPS 2023 Workshop on XAI in Action: Past, Present, and Future Applications.

DIGITAL

SLR: Explainable Artificial Intelligence (XAI) in (Financial) Time Series Forecasting

Reil, J. et al.

DIGITAL

Funded by
the European Union

Finance, just like other fields, is increasingly data-driven…

The stock market is not as it used to be…

The stock market evolved from in-person traders to computerized networks

DIGITAL

Finance, just like other fields, is increasingly data-driven…

The stock market is not as it used to be…

The stock market evolved from in-person traders to computerized networks

Movement into data-driven finance results in…

A movement into a field that is data-rich, but thus more complex by the day

DIGITAL

Finance, just like other fields, is increasingly data-driven…

The stock market is not as it used to be...

The stock market evolved from in-person traders to computerized networks

Movement into data-driven finance results in…

A movement into a field that is data-rich, but thus more complex by the day

Financial forecasting also gets impacted by the new trend…

Forecasting was relying on traditional models that were *used* to provide accurate and transparent predictions for forecasts

DIGITAL

Finance, just like other fields, is increasingly data-driven…

The stock market is not as it used to be...

The stock market evolved from in-person traders to computerized networks

Movement into data-driven finance results in…

A movement into a field that is data-rich, but thus more complex by the day

Financial forecasting also gets impacted by the new trend…

Forecasting was relying on traditional models that were *used* to provide accurate and transparent predictions for forecasts

AI approaches such as ML and DL have emerged…

These "black-box" methods have a promise of higher predictive accuracy, but at the cost of increased complexity

DIGITAL

Finance, just like other fields, is increasingly data-driven…

The stock market is not as it used to be...

The stock market evolved from in-person traders to computerized networks

Movement into data-driven finance results in…

A movement into a field that is data-rich, but thus more complex by the day

Financial forecasting also gets impacted by the new trend…

Forecasting was relying on traditional models that were *used* to provide accurate and transparent predictions for forecasts

AI approaches such as ML and DL have emerged…

These "black-box" methods have a promise of higher predictive accuracy, but at the cost of increased complexity

XAI (Explainable Artificial Intelligence) may be the key…

XAI seeks to make complex model behavior (like ML and DL) transparent and comprehensive to human users

DIGITAL

Finance, just like other fields, is increasingly data-driven…

The stock market is not as it used to be...

The stock market evolved from in-person traders to computerized networks

Movement into data-driven finance results in…

A movement into a field that is data-rich, but thus more complex by the day

”In finance, **an accurate prediction is useless if no one can explain it**”

Financial forecasting al
impacted by the new tre

ng on traditional models
to provide accurate and
predictions for forecasts

AI approaches such as ML and DL have emerged…

These “black-box” methods have a promise of higher predictive accuracy, but at the cost of increased complexity

XAI (Explainable Artificial Intelligence) may be the key…

XAI seeks to make complex model behavior (like ML and DL) transparent and comprehensive to human users

DIGITAL

Motivation

- ML and DL may improve financial forecasting predictions

- These ML and DL methods are mostly “black boxes”. Meaning: low transparency

- Financial institutions require trustworthy, explainable, and regulatory-compliant AI (e.g., GDPR)

- **XAI** (Explainable AI) could bridge performance and interpretability (*best of both worlds*) but **remains underdeveloped for time series**

Problem

Despite the rapid progress of artificial intelligence (AI) in financial time series forecasting, existing explainable AI (XAI) approaches remain fundamentally misaligned with the statistical properties of financial data.

DIGITAL

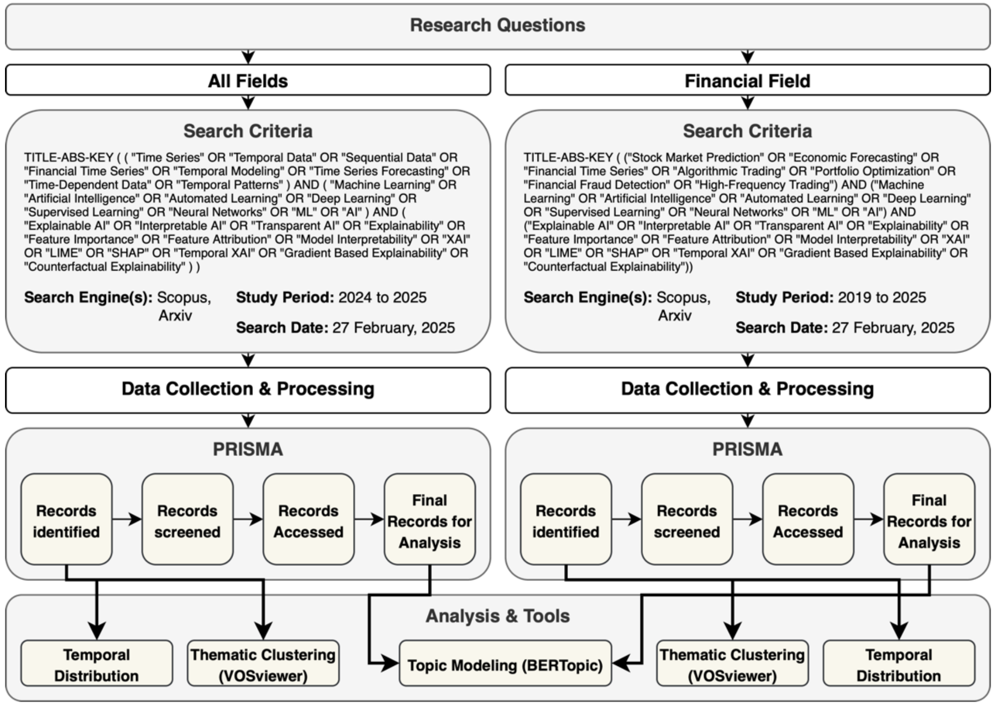
Methodology

Title: Explainable Artificial Intelligence (XAI) in Time Series Forecasting: Insights from the Literature

Focus: Literature review on temporal data, ML, and XAI to identify gaps and guides for future research

Methodology:

1. Search academic databases (Scopus, arXiv) for studies on XAI + time series forecasting
2. Include both financial and non-financial applications to capture methodological width
3. Perform screening and inclusion using defined keywords and eligibility criteria
4. Extract and select data on (X)AI methods, validation practices, and limitations
5. Apply bibliometric and topic-modeling analysis (e.g., VOSviewer, BERTopic) to identify clusters and trends in papers
6. Synthesize results into an overview of XAI approaches for time series data
7. Identify research gaps in method design and evaluation



DIGITAL

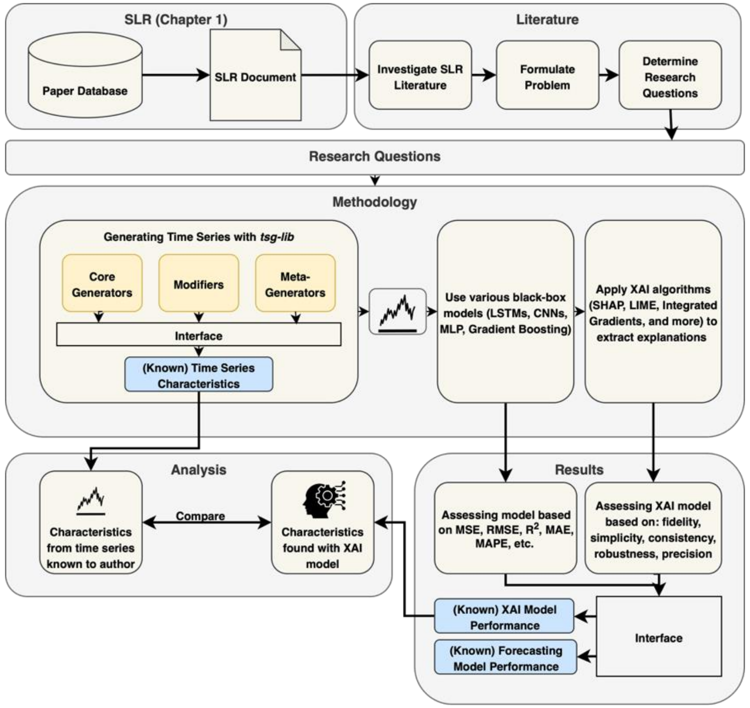
Methodology

Title: Evaluating XAI Methods for Financial Time Series Forecasting with Synthetic & Real Data

Focus: Empirically test how statistical properties of financial time series affect XAI methods

Methodology:

1. Develop a controlled experimental framework using both synthetic and real financial time series
2. Generate synthetic datasets with predefined statistical properties (e.g., non-stationarity, volatility, autocorrelation) using *tsg-lib*
3. Select deep learning forecasting models (LSTM, CNN) and baseline econometric models
4. Apply multiple XAI techniques (SHAP, LIME, Integrated Gradients, Grad-CAM) across models
5. Evaluate explanations with quantitative explainability validation metrics such as fidelity, stability, consistency, robustness
6. Compare model–XAI performance across different statistical conditions



DIGITAL

Literature Review

What	A SLR was conducted within the field of temporal data, ML, and XAI
Method	PRISMA

Investigation:	Financial Time Series
	AI for Financial Time Series
	XAI for Financial Time Series
	Explainability Validation Metrics
	Limitations/Challenges of Current XAI Methods

DIGITAL

Literature Review

What SLR is broad, incorporating various aspects (such as statistical properties of time series)

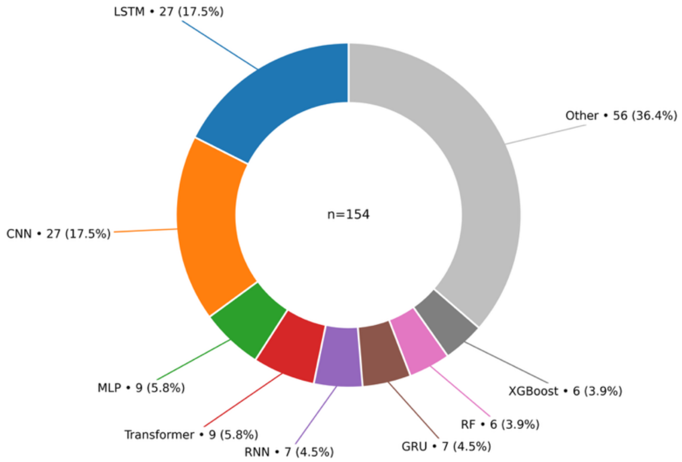
Finding Financial data are non-linear, evolving, and dependent over time (properties fit all to a certain degree)

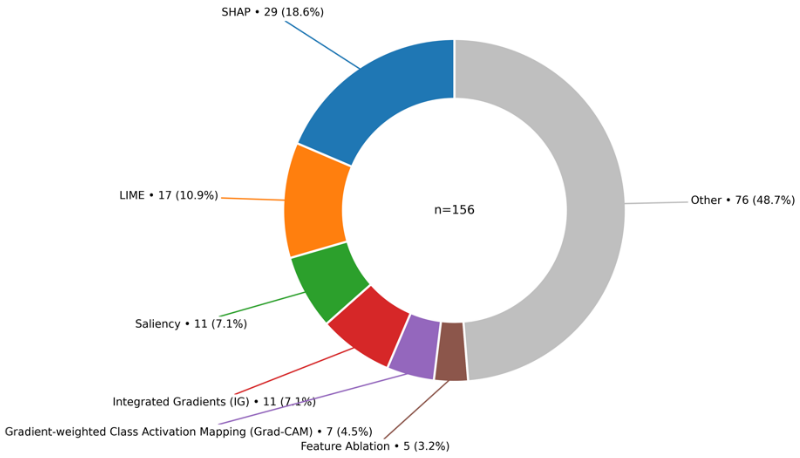
Statistical property	Explanation	Reference
Temporal dependencies (sequential nature)	Refers to the relationship between financial time series' time points.	Iqbal et al. (2025)
Time-varying dependence structure	Refers to a non-constant joint distribution of variables over time.	Cont (2001)
Non-stationarity	Indicates that the statistical properties of the financial time series, such as mean and variance, change over time.	Cai et al. (2024)
Trend, seasonality, irregularity (noise and residuals)	Represent the systematic components into which a financial time series can be decomposed.	Chakraborty et al. (2024)
Autocorrelation	Measures the linear relationship between a financial time series and its lagged values, capturing how past observations influence current ones.	Naser and Naser (2025)
Heteroskedasticity (volatility clustering)	Describes time-varying variance in financial time series, where periods of high volatility tend to cluster together.	Bollerslev (1986) and Engle (1982)
Heavy tails (fat-tailedness)	Refers to the empirical observation that return distributions often exhibit higher kurtosis than the normal distribution.	Cont (2001)
Regime shifts / structural breaks	Sudden changes in the financial time series due to macroeconomic events, policy changes, or crises.	Perron (2006)

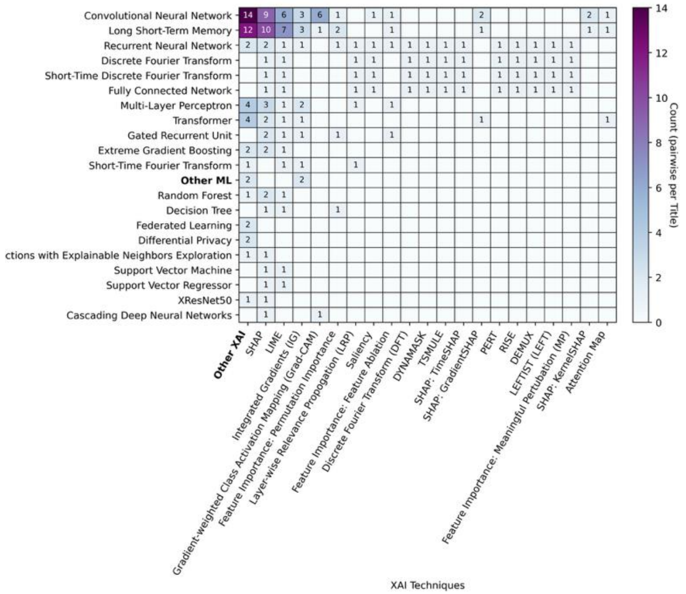
DIGITAL

Literature Review

What An investigation was also conducted into (X)AI for financial time series and its limitations







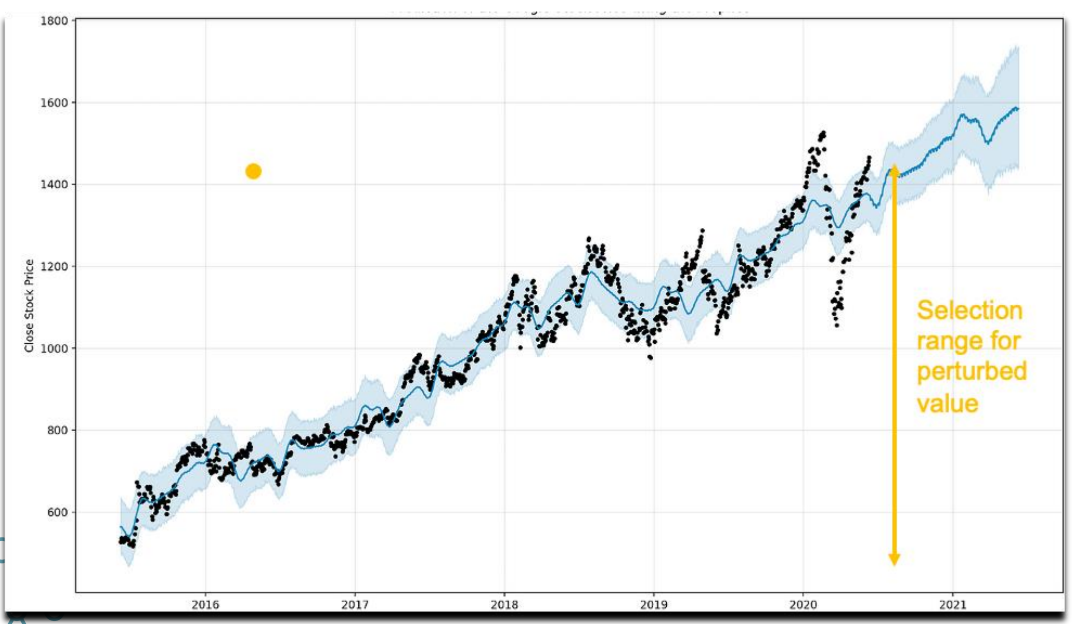
DIGITAL

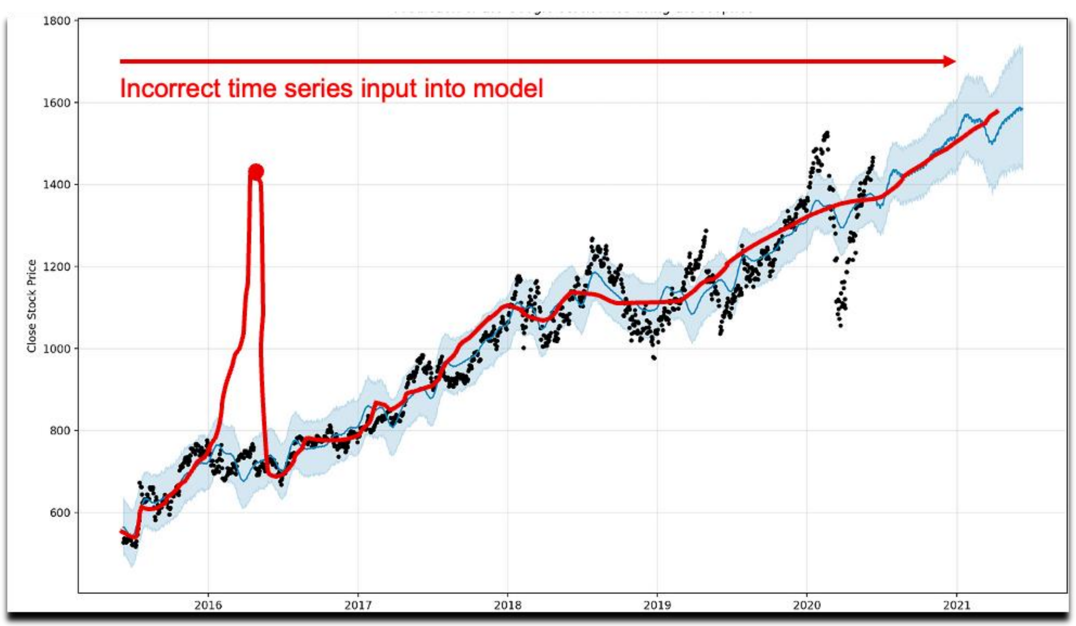
(Way more explanations have been generated in the SLR to support findings made…)

Literature Review

Finding The investigation revealed that existing XAI often fails to produce stable and faithful explanations due to the statistical properties of financial time series

Example: perturbation for XAI in time series (e.g., SHAP)





DIGITAL

Literature Review

How It was necessary to also investigate how XAI methods were evaluated

Finding Literature found that XAI methods not always were as well validated as they should, however, explainability validation metrics were used in cases

Explainability validation metric	Explanation	Reference(s)
Fidelity (faithfulness, perturbation analysis)	Measures how accurately an explanation reflects the model's true decision-making process.	Theissler et al. (2022)
Stability	Assesses the consistency of explanations under small input perturbations.	Misheva and Osterrieder (2023)
Consistency	Evaluates whether different XAI methods or model instances produce similar explanations for similar inputs.	Ndao et al. (2025)
Robustness	Examines the resilience of explanations to adversarial or structured perturbations in the data or model.	Brusch et al. (2024)

Explainability validation metrics are a full field that is involving… (Pawlicki et al., 2024)

DIGITAL

Research Gaps

"Despite progress in AI-based forecasting, **existing XAI approaches are misaligned with** the statistical properties of **financial data**"

Standard XAI assumes independent, stable inputs

Financial data violate these assumptions

Explanations can be inconsistent, unstable, or wrong

and more...

Technical Gap
Current XAI methods fail to account for the statistical properties of financial time series

Human-centric Gap
Explanations are not designed around the interpretive needs of financial stakeholders and their decision contexts

DIGITAL

Latest advancements in explainable artificial intelligence

Altukhi, Z. M., Pradhan, S., & Aljohani, N. (2025).
A systematic literature review of the latest advancements in explainable artificial intelligence.
Technologies, 13(3), 93. https://doi.org/10.3390/technologies13030093

DIGITAL

Funded by the European Union

Latest Advancements

- Altukhi et al. (2025) provide a PRISMA-based systematic review of XAI research from 2014–2024, categorizing recent advances into model developments, evaluation methods, and user-centered XAI systems, and highlighting a shift toward bridging technical explainability and human understanding.

- Purpose → To systematically map recent progress in XAI and identify dominant research directions and gaps.

- Methodology
  - PRISMA-based systematic literature review
  - Sources: IEEE Xplore, ACM, ScienceDirect
  - Time span: 2014–2024
  - Final selection: 30 key studies

DIGITAL

Latest Advancements

Figure 5. XAI advancements categories across the 30 articles (five studies cover more than one category, as shown in Table 4).

DIGITAL

Theme 1: Model development

- **Focus**
  - New algorithms and frameworks to explain black-box models
  - Both **model-specific** and **model-agnostic** approaches

- **Examples**
  - Feature attribution (SHAP, LIME, gradients)
  - Rule-based and surrogate models
  - Architecture-aware explanations

- **Key insight**
  - Technical sophistication has increased
  - But **interpretability ≠ usability**
  - *Explanations often remain mathematically correct but cognitively opaque.*

DIGITAL

Theme 2: Evaluation Metrics & Methods (How Good Are Explanations?)

- **What is evaluated**
    - Faithfulness
    - Robustness
    - Stability
    - Completeness
- **Challenges identified**
    - No **standardized evaluation benchmarks**
    - Heavy reliance on **proxy metrics**
    - Limited agreement on "ground truth explanations"
- **Key finding**
    - Evaluation of XAI methods is **fragmented and inconsistent**, making comparisons difficult.
        - This theme directly motivates work like **AttributionLab** and **faithfulness studies**.

DIGITAL

Theme 3: User-Centered XAI Systems (Who Are Explanations For?)

- **Shift in focus**
  - From *model inspection → human understanding*
  - Different stakeholders:
    - Developers
    - Domain experts
    - Non-technical users
- **Key directions**
  - Interactive XAI systems
  - Personalized explanations
  - Domain-adaptive interfaces
- **Main insight**
  - Effective XAI must be designed as a **human–AI communication system**, not just a diagnostic tool.
  - Strong overlap with **TalkToModel** and application-grounded evaluation.

DIGITAL

Theme 4: Open Challenges

- **Open challenges**
  - Lack of unified evaluation standards
  - Weak connection between explanations and decision-making
  - Limited deployment-focused XAI (monitoring, lifecycle)

- **Core takeaway**
  - The future of XAI lies not in better explanations alone, but in **better alignment between models, humans, and real-world use**.

Towards Explanatory Model Monitoring

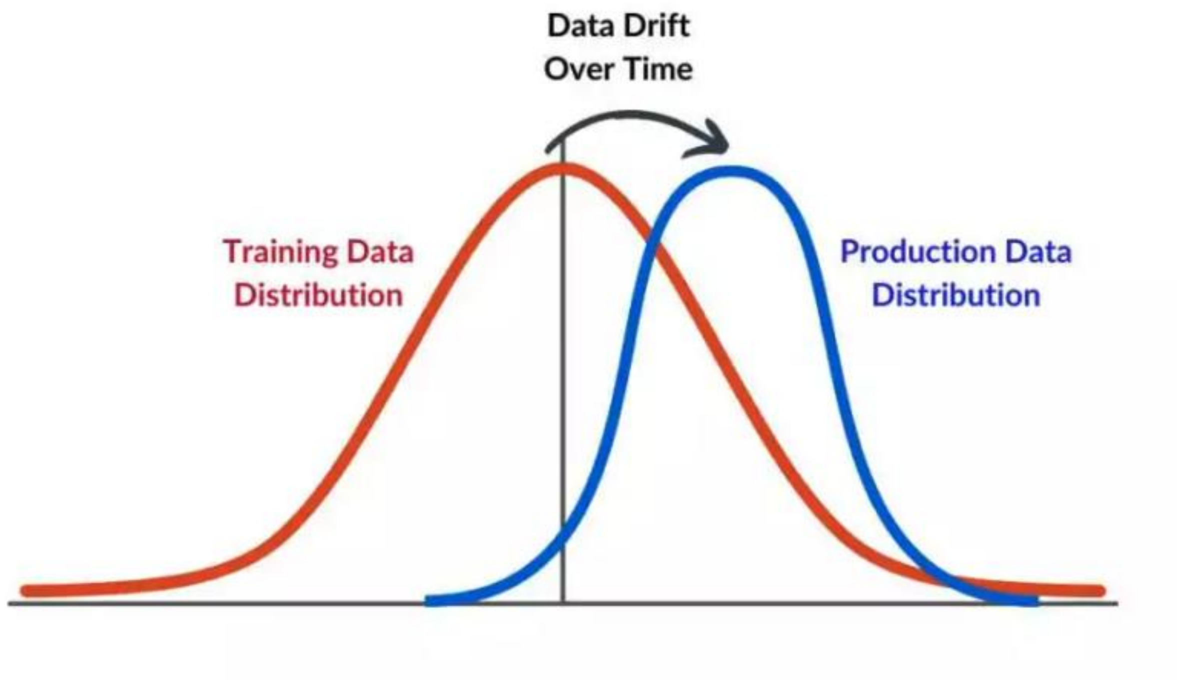
Koebler, A., Decker, T., Lebacher, M., Thon, I., Tresp, V., & Buettner, F. (2023). Towards explanatory model monitoring. Proceedings of the NeurIPS 2023 Workshop on XAI in Action: Past, Present, and Future Applications.

DIGITAL

Funded by
the European Union

Towards Explanatory Model Monitoring

- In real-world ML systems:
  - Models degrade over time due to data drift, sensor issues, pipeline changes, or real-world changes.
  - Traditional monitoring:
    - Performance monitoring needs labels (often unavailable or delayed).
    - Drift detection flags changes but does not explain whether they matter or what to do.



Practitioners are left asking:

"The model got worse - but which features caused it, and what should we fix?"

DIGITAL

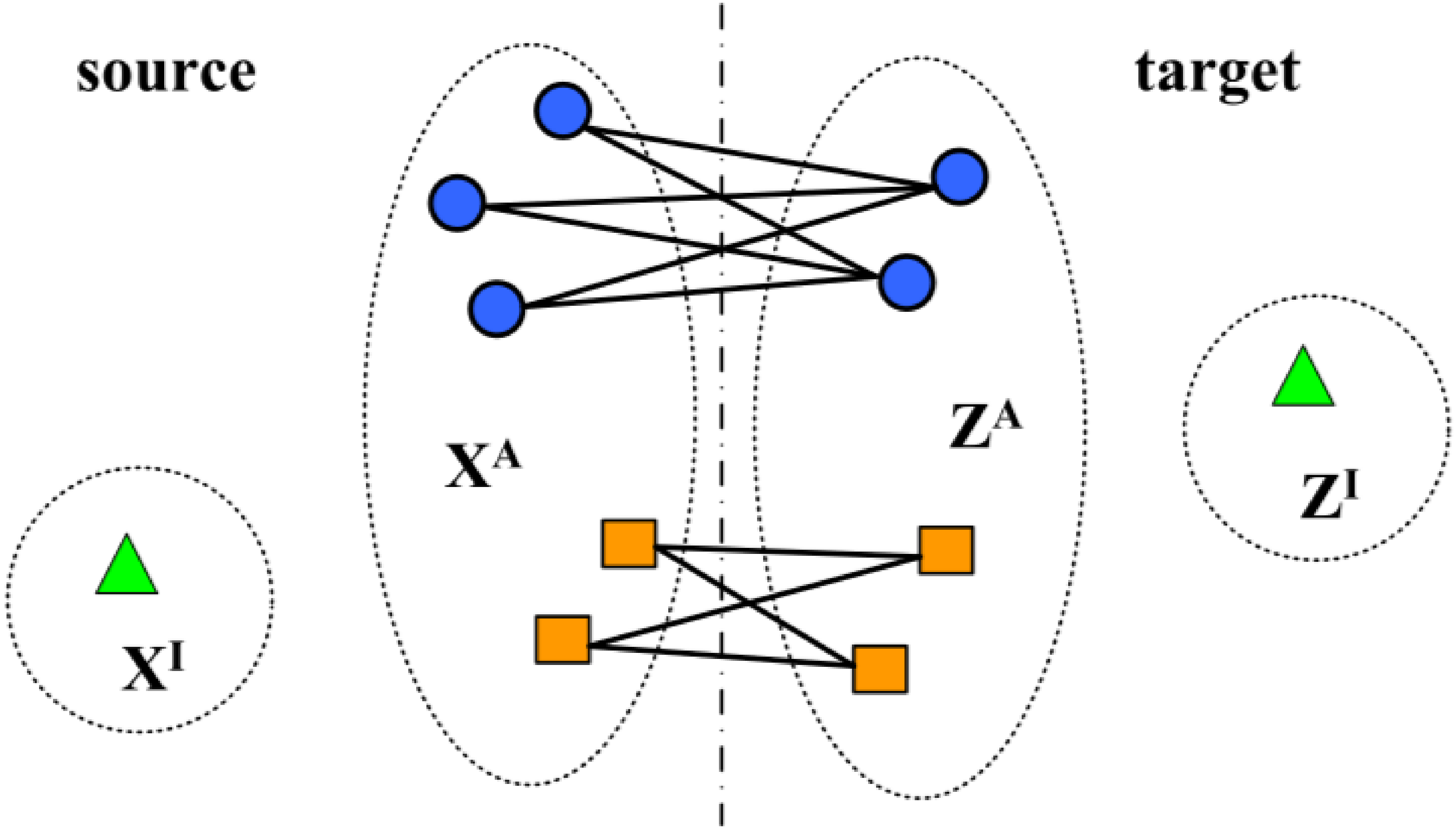
Towards Explanatory Model Monitoring

- Main contribution: **Explanatory Performance Estimation (XPE)**
- The paper introduces XPE → Estimates how much model performance has changed and attributes that change to specific input features - without target labels.

- So XPE answers two questions simultaneously:
  - How much performance likely dropped?
  - Which features are responsible for that drop?

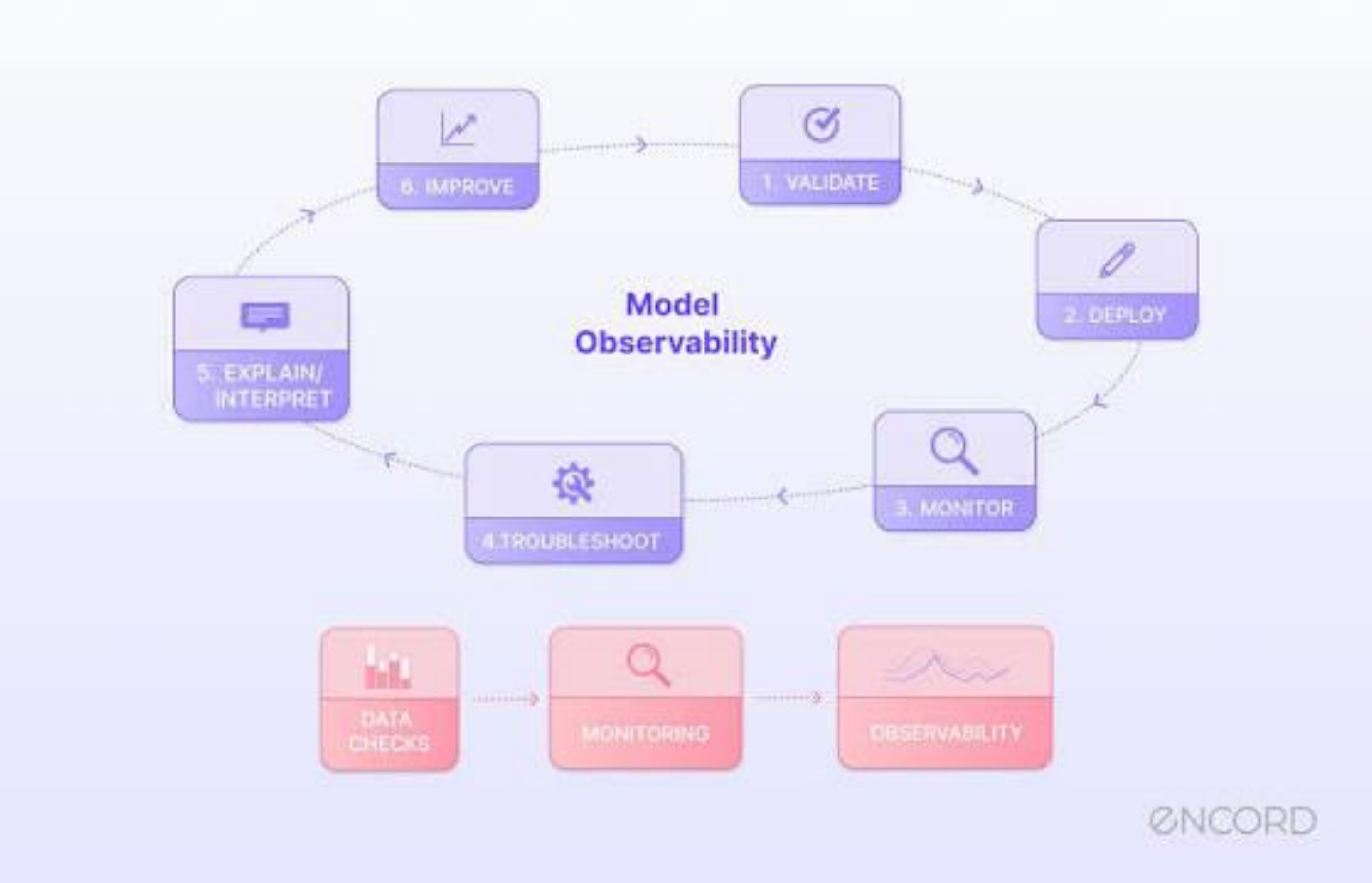
DIGITAL

Towards Explanatory Model Monitoring



DIGITAL

Towards Explanatory Model Monitoring



DIGITAL

Key idea (**intuition**)

- **Assume a distribution shift** from training data $P_s(X, Y)$ to deployment data $P_t(X)$
- **No labels are available** at deployment

- Instead of:
  - just measuring drift, or
  - just tracking attribution scores,

→ XPE **connects drift, predictions, and explainability**.

- **How?**
  - It **aligns source and target data** using *optimal transport*
  - Compares **model behavior before vs. after the shift**
  - Uses **feature attributions** to decompose the *estimated performance change* into **feature-level contributions**

DIGITAL

What makes this **different from prior work**?

Existing approach	Limitation
Performance monitoring	Needs labels
Drift detection	No link to performance
Attribution drift tracking	Importance can change even if performance doesn't
Causal/XAI methods	Require full causal knowledge

XPE uniquely provides:

- Label-free performance estimation
- Feature-level explanations of degradation
- Actionable diagnostics (e.g. sensor failure vs. concept drift)

DIGITAL

What you get as an output?

- Instead of just an alert like:

"Data drift detected"

- You get something like:

Estimated accuracy dropped by 8%
70% of the drop is attributable to features X_1 and X_3
Likely root cause: sensor malfunction on X_1

- This is why the authors call it **actionable model monitoring**.

DIGITAL

DIGITAL

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Horizon Europe: Marie Skłodowska-Curie Actions. Neither the European Union nor the granting authority can be held responsible for them.

DIGITAL

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101119635