The Rise of Tabular Foundation Models:

From Zero-Shot Inference to the Security frontier

17.02.2026 | Data Science Meetup Luxembourg

Karim Tit

UNIVERSITY OF LUXEMBOURG

Who I am: Karim Tit



Current position (2024-):

Postdoctoral Researcher in AI Safety at SnT

Research Interests:

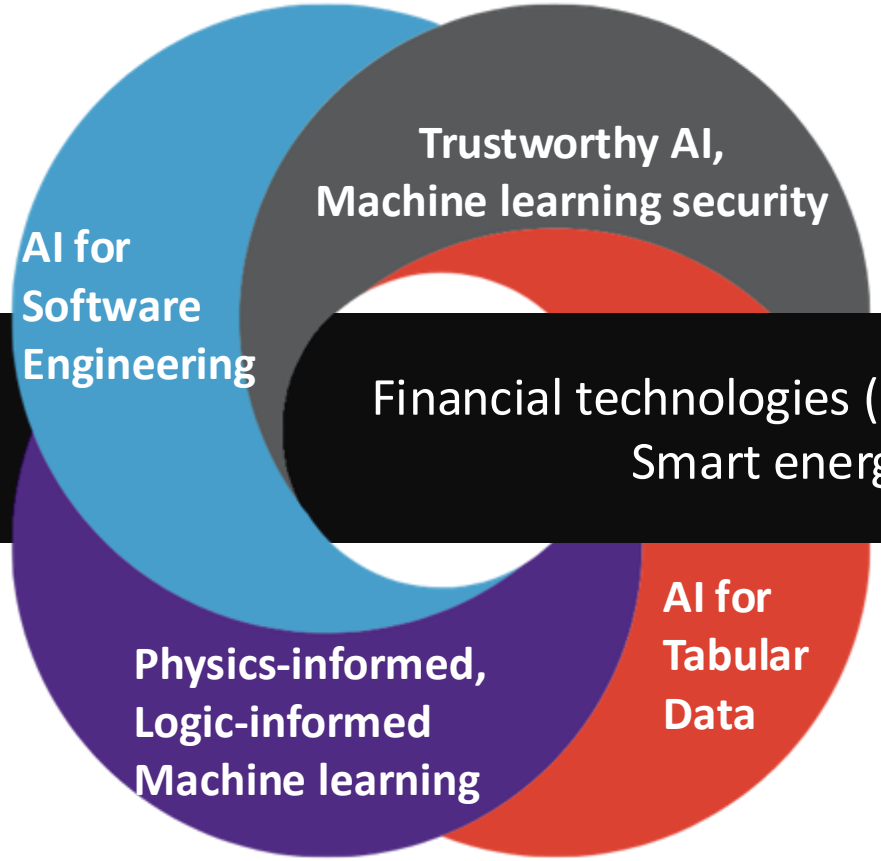
- Empirical and Certified Robustness of Foundation Models
- Logics-informed and Constraints-Informed Generative AI

Previous position (2021-2024):

Industry PhD Student @ Thales and @ Inria

Subject: "Reliability of Deep Learning with Rare Event Simulation"

Who we are: Security, Reasoning and Validation (SerVal) research group



Financial technologies (Paypal, BGL BNP Paribas), Smart manufacturing (Cebi, Ceratizit)
Smart energy systems (Creos, Enovos), Public services (STATEC)

On the Menu Today:

1. An Introduction to Tabular Foundation Models (TFMs)
2. A Zoom on TabPFN: its Architecture and *Philosophy*
3. A Tour of State-of-the-Art TFMs & Scalability Issues
4. A Security Perspective: Evasion Attacks on TFMs (SatML'26 paper)
5. Conclusion & QA session

SNT

The Rise of Tabular Foundation Models



Summary

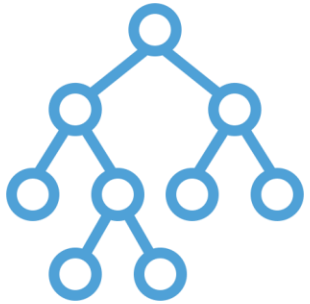
Introduction

A Zoom on TabPFN

UNIVERSITY OF LUXEMBOURG

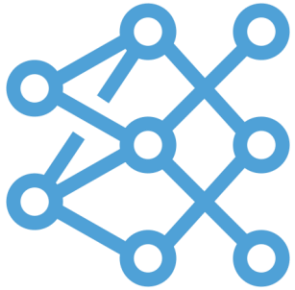
Background : Tabular Machine Learning History

A new paradigm shift ?



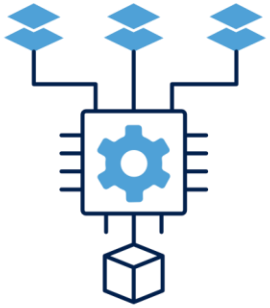
Tree-Based Models Domination (≈2000s–2021)

- Strong inductive bias
- *Interpretability*
- **GBDTs Dominates** Tab ML, particularly competitive DS



Neural Network Resurgence (≈2021–2024)

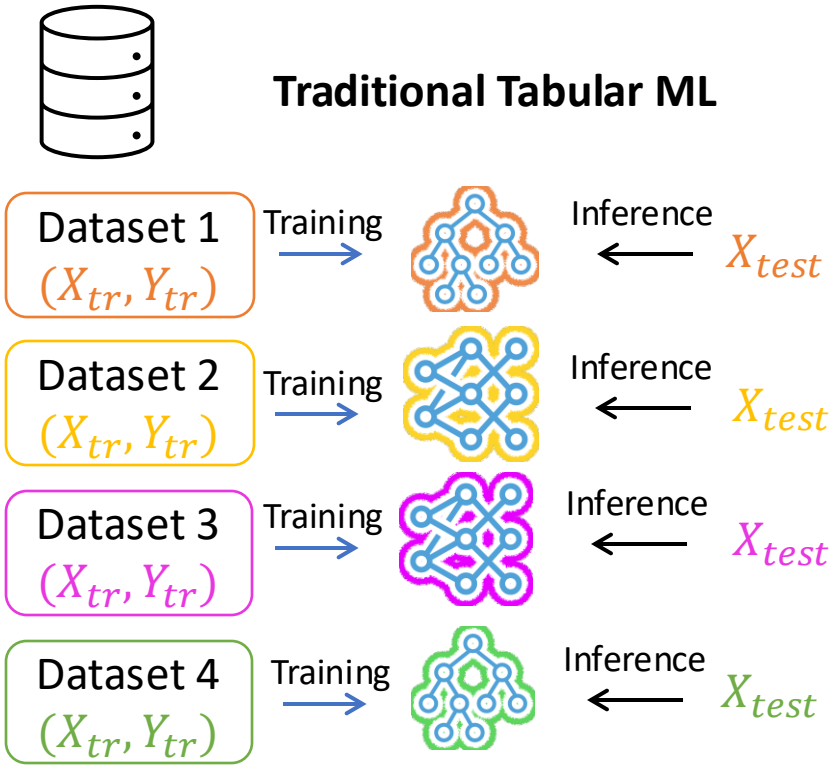
- Modalities Interactions
- **Strong performance** on very **large-scale** data
- Still dominated by GBDTs in small/medium data regimes

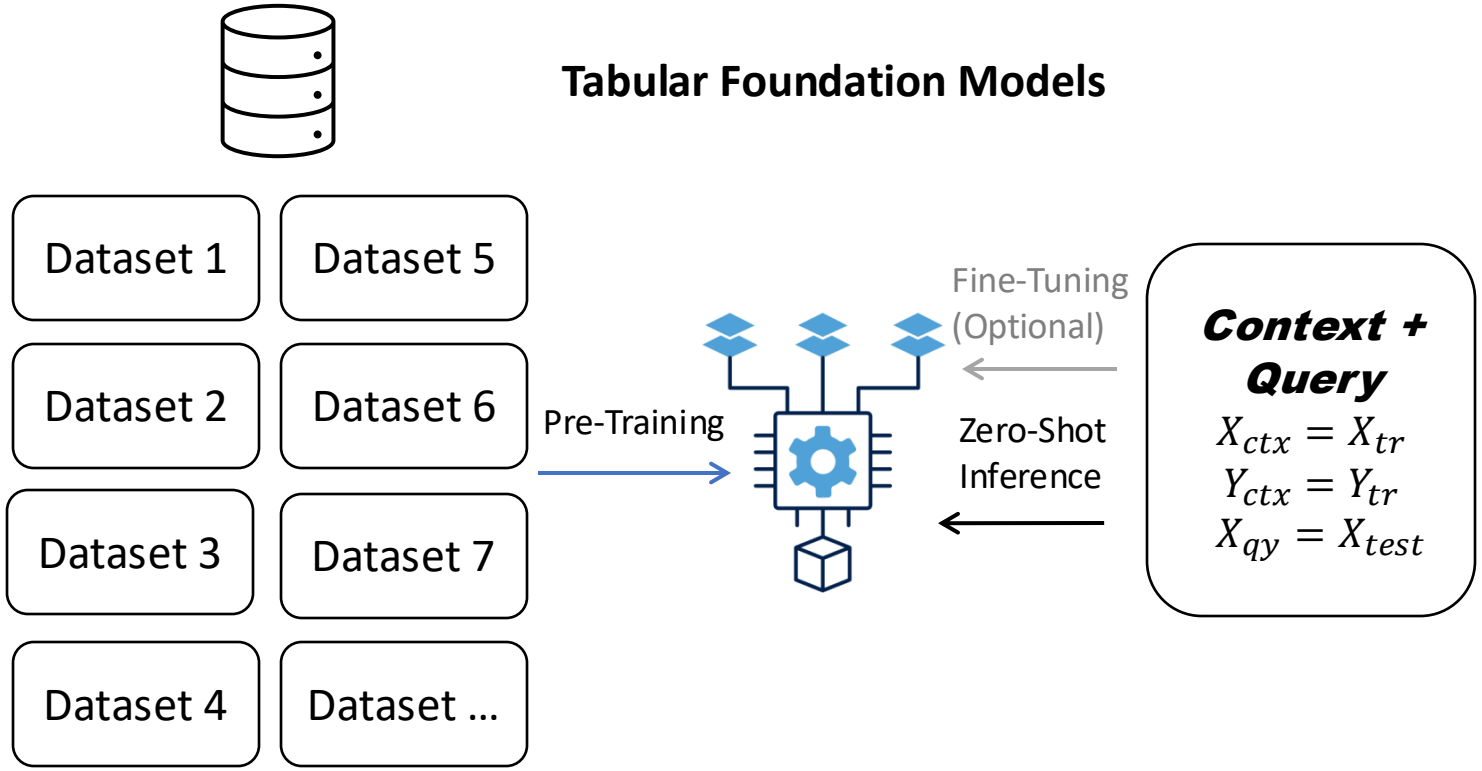


Tabular Foundation Model Era ? (≈2024–Present)

- Pre-trained on Large Synthetic tabular "dataset of datasets"
- **In-Context Learning** focus
- Dominant on **small-scale** data

What are Tabular Foundation Models ?





In Traditional ML the focus is on **Model Selection**:
→ Choosing the right model for a given dataset

With TFMs the focus moves to **Dataset and Architecture Selection**:
1. Choosing the right datasets for the **Pre-Training** phase
2. Choosing the right **global** architecture for **scalable** Inference

So, what are the main TFMs on the block ?

- **TabPFN** (PriorLabs): Meta-trained transformer performing amortized Bayesian inference via in-context learning.
- **RealTabPFN** (Prior Labs): TabPFN extended with real-data pretraining to improve downstream accuracy.

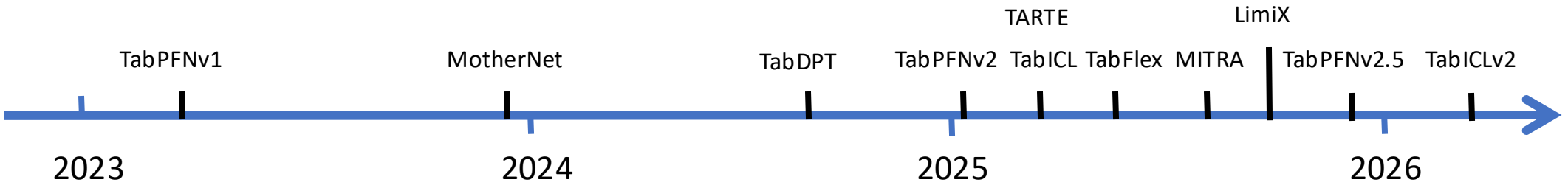
- **TabICL** (Inria): Scalable in-context transformer using distribution-aware **column compression**.
- **TARTE** (Inria): Semantics-aware table encoder leveraging **column names** and **textual metadata**.

- **MotherNet** (Microsoft): Pretrained hypernetwork designed for rapid adaptation to new tabular tasks.
- **TabFlex** (Microsoft): Linear-attention PFN variant enabling efficient large-scale in-context inference.

- **MITRA** (Amazon): PFN-style model pretrained on a **curated and diverse mixture** of synthetic data generators.

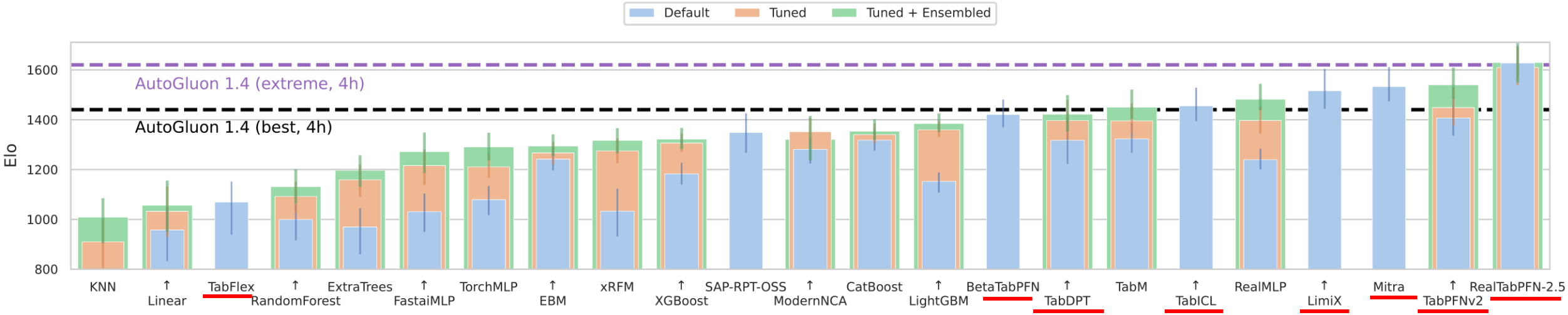
- **TabDPT** (Layer 6 AI): In-context learner with **column-masking self-supervised** pretraining on **real data**.

- **LimiX** (Stable AI / Tsinghua): Unified masked-distribution model supporting tabular prediction, **imputation**, and generation.



How "good" really are TFMs ?

On "small" datasets of TabArena (<10K samples, <500 features):

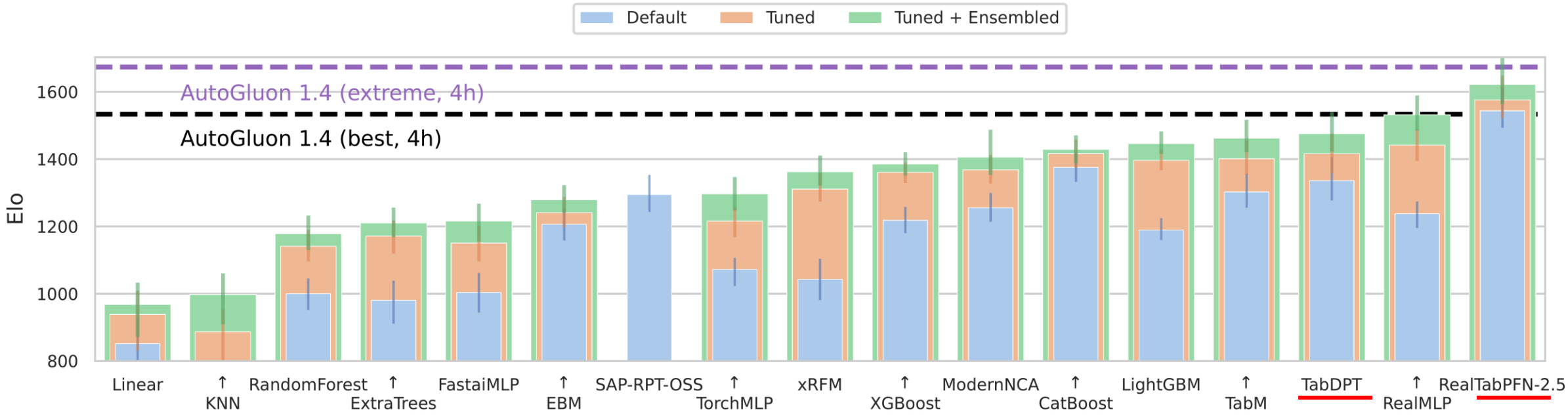


TabArena Elo Rankings Leaderboard for 26 datasets (TabPFNv2-compatible datasets, classification tasks) including all models. TabPFNv2-compatible datasets contain at most 10,000 samples, 500 features, and 10 classes. (As of February 13th, 2026)

Nick Erickson et al. "TabArena: A Living Benchmark for Machine Learning on Tabular Data", Preprint., 2025.

How "good" really are TFMs ?

On the Entire TabArena Benchmark, a wide range of 51 datasets:



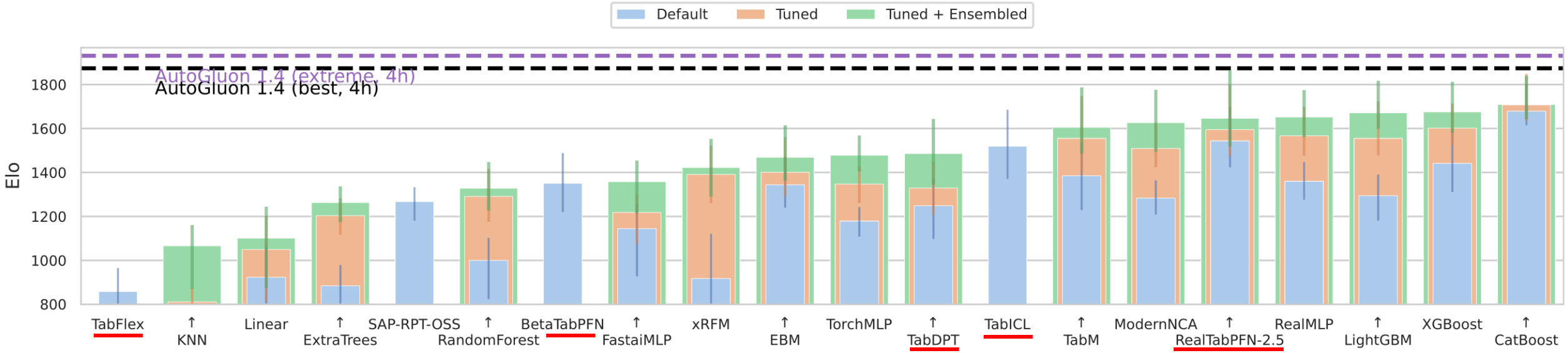
TabArena Elo Rankings Leaderboard for 38 datasets (all datasets, classification tasks) including all models.

(As of February 13th, 2026).

Nick Erickson et al. "TabArena: A Living Benchmark for Machine Learning on Tabular Data", Preprint., 2025.

How "good" really are TFMs ?

On "medium" datasets of TabArena (containing between 10K and 250K samples):



TabArena Elo Rankings Leaderboard for 10 datasets (medium datasets, classification tasks) including all models.

Medium datasets contain between 10,000 and 250,000 samples.

(As of February 13th, 2026)

Nick Erickson et al. "TabArena: A Living Benchmark for Machine Learning on Tabular Data", Preprint., 2025.

SnT

The Rise of Tabular Foundation Models



Introduction

A Zoom on TabPFN

A Tour of State-of-the-Art TFMs

uni.lu
UNIVERSITY OF LUXEMBOURG

Background: Impact of TabPFN

PRIOR LABS

Widespread Data Science Community Adoption

400
Citations

2 000 000 +
PyPI

1500
Users

Real-World Impact Through Applied Research

Healthcare & Life Sciences (50+)

Energy & Utilities (15+)

Manufacturing & Industrial (12+)

Other Sectors (19+)

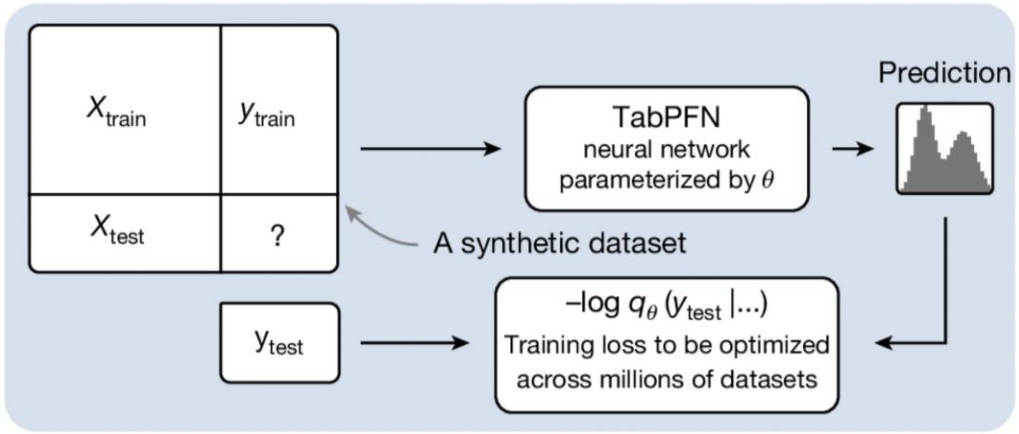
Grinsztajn, Léo, et al. "TabPFN-2.5: Advancing the State of the Art in Tabular Foundation Models." *arXiv preprint arXiv:2511.08667* (2025).

What are Prior-Fitted Networks ?

The TabPFN Model as a Canon

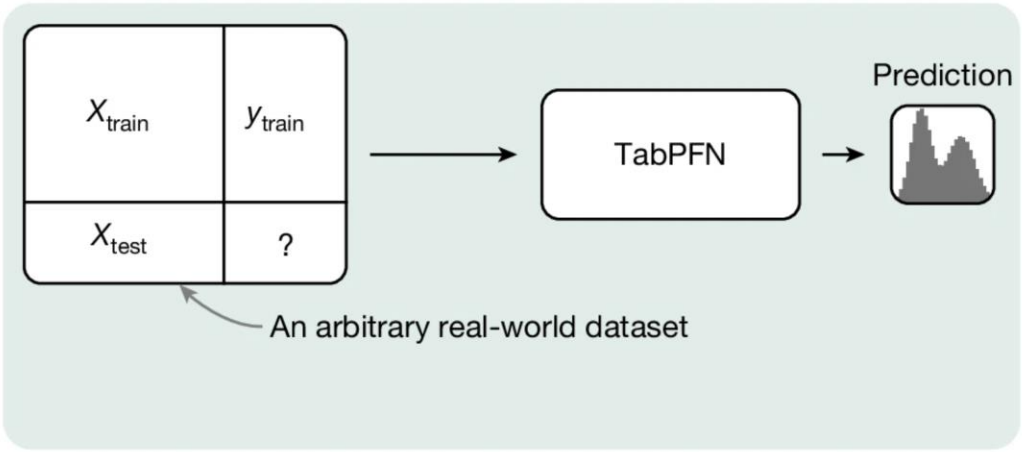
1) Pretraining

TabPFN is trained on synthetic data to take **entire datasets** as inputs and predict test labels in a **single forward pass**



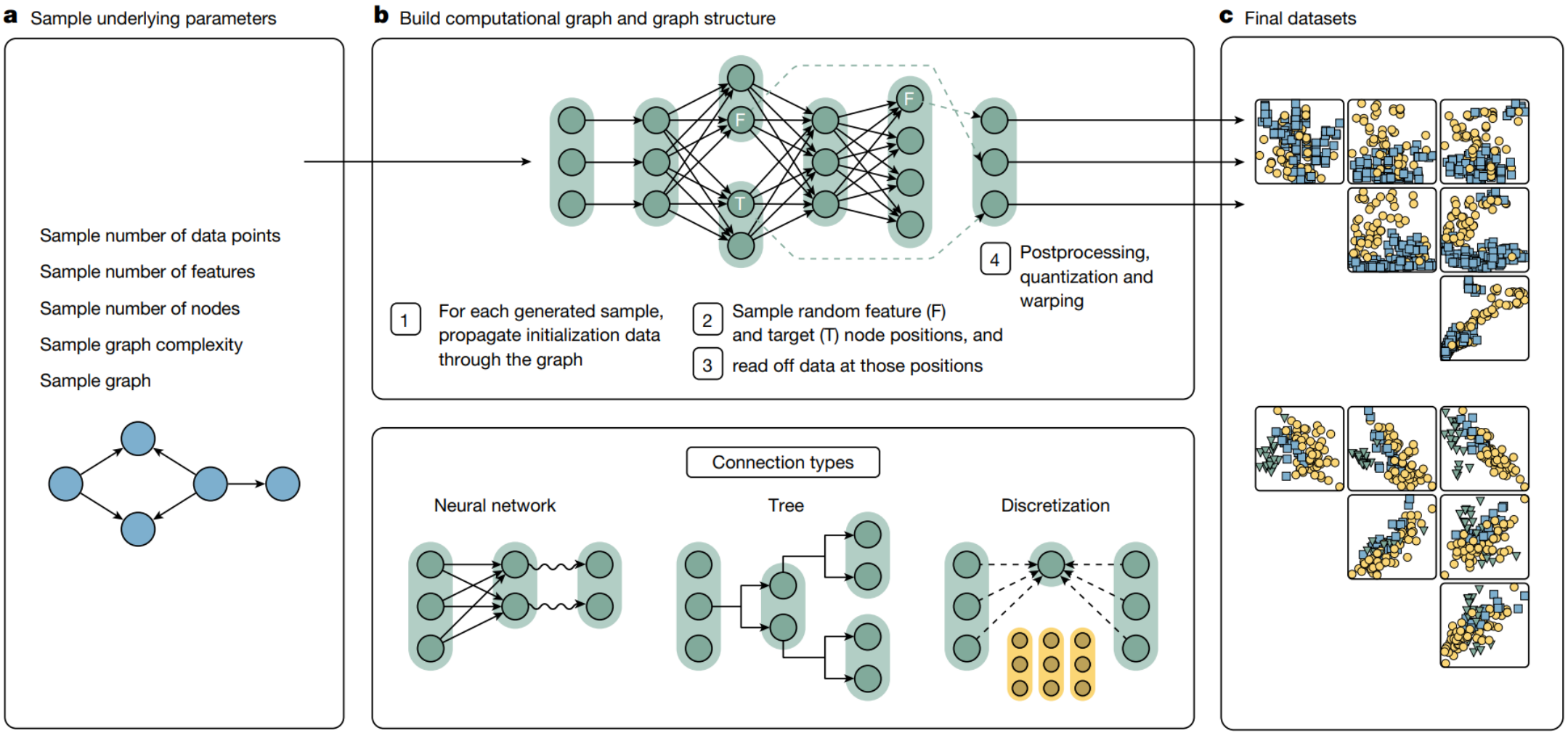
2) Prediction

TabPFN can now be applied to arbitrary **unseen** real-world datasets



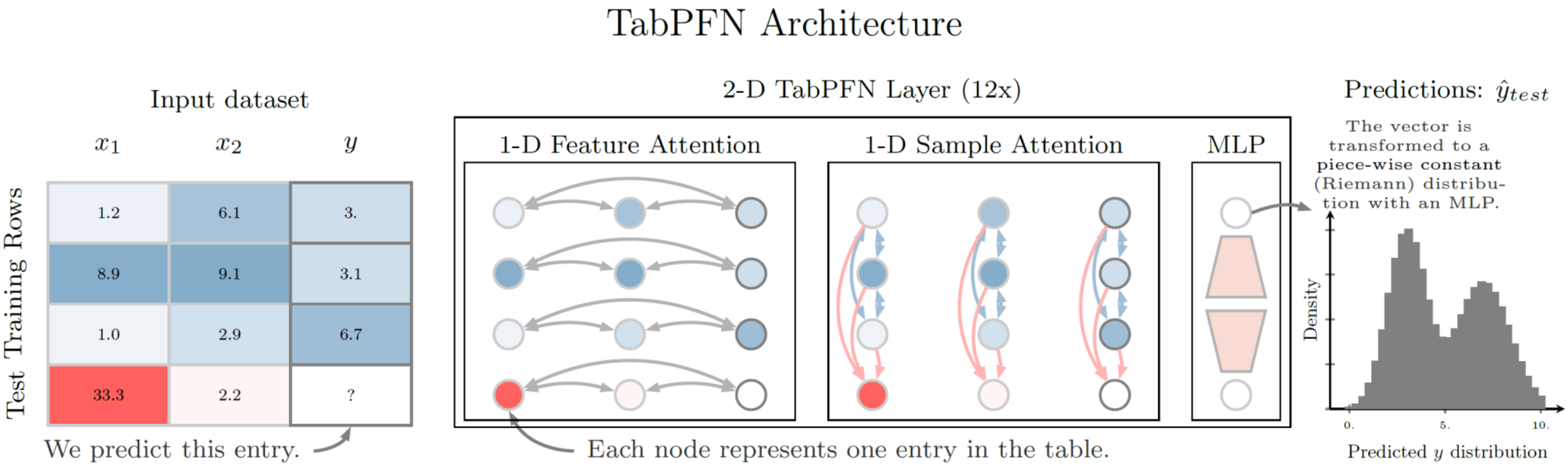
Hollmann, Noah, et al. "Accurate predictions on small data with a tabular foundation model." *Nature* 637.8045 (2025): 319-326.

Synthetic Data Generation Process



Hollmann, Noah, et al. "Accurate predictions on small data with a tabular foundation model." *Nature* 637.8045 (2025): 319-326.

TabPFNv2's Transformer Architecture



Hollmann, Noah, et al. "Accurate predictions on small data with a tabular foundation model." *Nature* 637.8045 (2025): 319-326.

TabPFN's Capabilities Evolution

Model	Rows	Feat.	Type	Depth	Inference mode
TabPFN-v1	1,000	100	Num.	8	ICL
TabPFN-v2	10,000	500	Mixed	12	ICL
TabPFN-2.5	50,000	2000	Mixed	18–24	ICL+MLP/Trees

Table 1: **Summary of TabPFN model variants.** Max Rows and Features are the recommended maximum sizes. The models also fit larger datasets but are not built and evaluated for these settings.

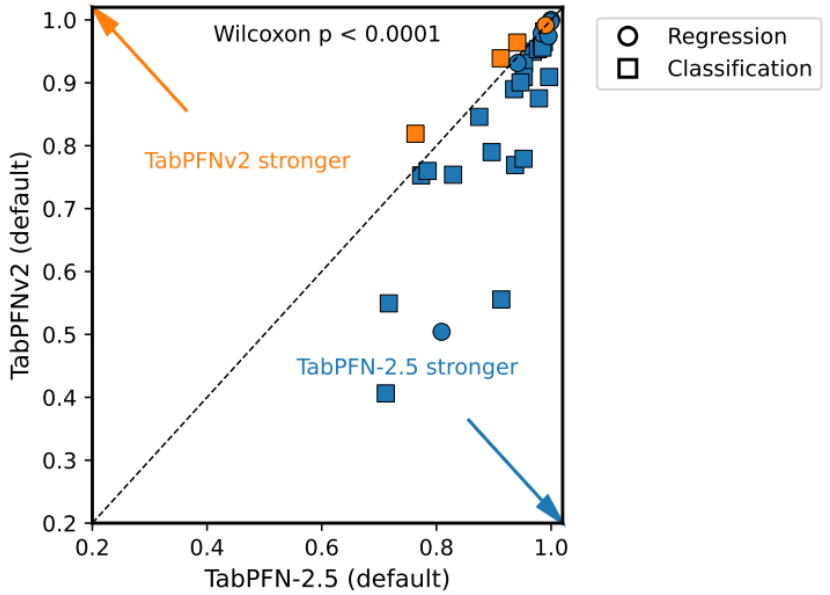


Figure 2: **TabPFN-2.5 clearly outperforms TabPFNv2.** We show normalized performance for each dataset of the TabPFNv2 subset of TabArena. TabPFN-2.5 often performs much better and is never much worse.

- The focus of TabPFN iterations has been **scalability**
- In addition, TabPFN versions performed consistently at or above SoTA

Grinsztajn, Léo, et al. "TabPFN-2.5: Advancing the State of the Art in Tabular Foundation Models." arXiv preprint arXiv:2511.08667 (2025).

Real TabPFN: Adding Real Data in the Training Mix

Core features

- Continued pre-training of TabPFNv2 on curated real-world datasets
- Two-stage training: synthetic pretraining → real-data adaptation
- L2 regularization to prevent catastrophic forgetting
- Significant improvement over vanilla TabPFNv2.5

Limitations

- Requires **careful** dataset curation
- Possible data-leaks in the evaluation process

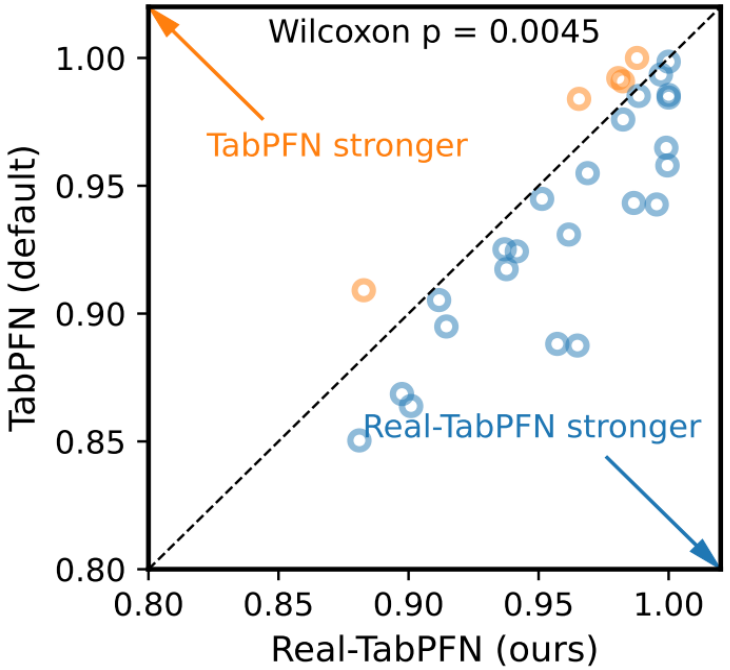


Figure 1. Per Dataset Normalized ROC Comparison of TabPFN (default) and Real-TabPFN (ours) on the 29 datasets from the OpenML AutoML Benchmark Datasets. Wilcoxon p refers to the two-sided Wilcoxon signed-rank test p value.

Garg, Anurag, et al. "Real-TabPFN: Improving Tabular Foundation Models via Continued Pre-training With Real-World Data." ICML Workshop on Foundation Models (2025).

SNT

The Rise of Tabular Foundation Models

A Zoom on TabPFN

A Tour of Alternative SOTA TFMs

Evasion Attacks on TFMs



UNIVERSITY OF LUXEMBOURG

TabICL

Core features

- Transformer-based tabular foundation model for in-context learning
- Column-wise inter-sample embeddings + row-level attention

v2 improvements

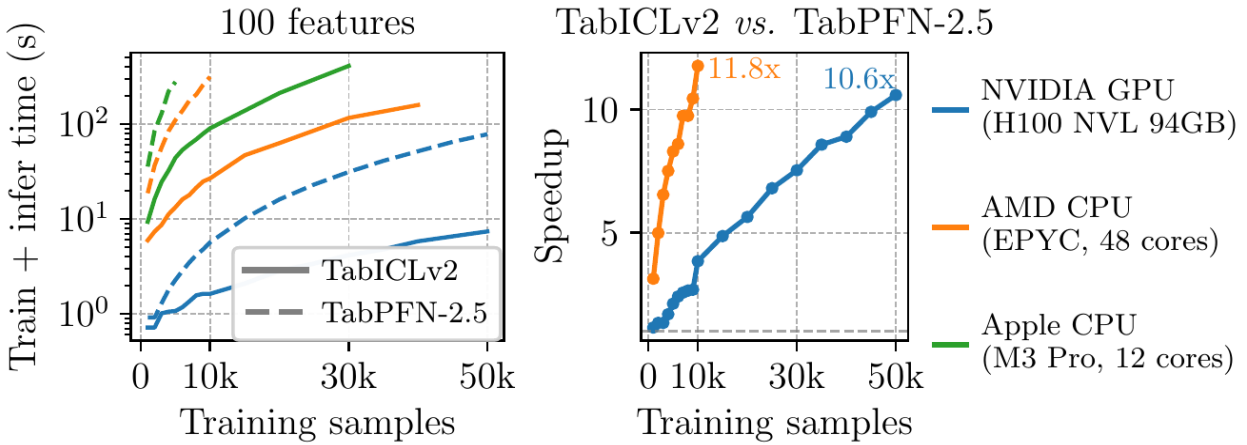
- Larger and more diverse pretraining regime
- Stronger performance across broader benchmarks

Strengths

- Scales to larger datasets than other PFN-style models
- Faster training **and** inference on both CPU and GPU
- **Fully open source pre-training procedure**

Limitations

- Still fully synthetic-pretrained
- No explicit prior-mixture optimization (unlike MITRA)



Qu, Jingang et al. "TabICL: A Tabular Foundation Model for In-Context Learning." ICML (2025).
Qu, Jingang et al. "TabICLv2: A better, faster, scalable, and open tabular foundation model" arxiv preprint, 2026

TARTE

Core features:

- Text-aware table encoder combining column names and cell values
- Joint encoding of numerical and string features

Strengths

- Leverages semantic information in feature names
- Works well when table **metadata is informative**

Limitations

- Less purely “in-context” than PFN-style models
- Performance depends on availability of column names

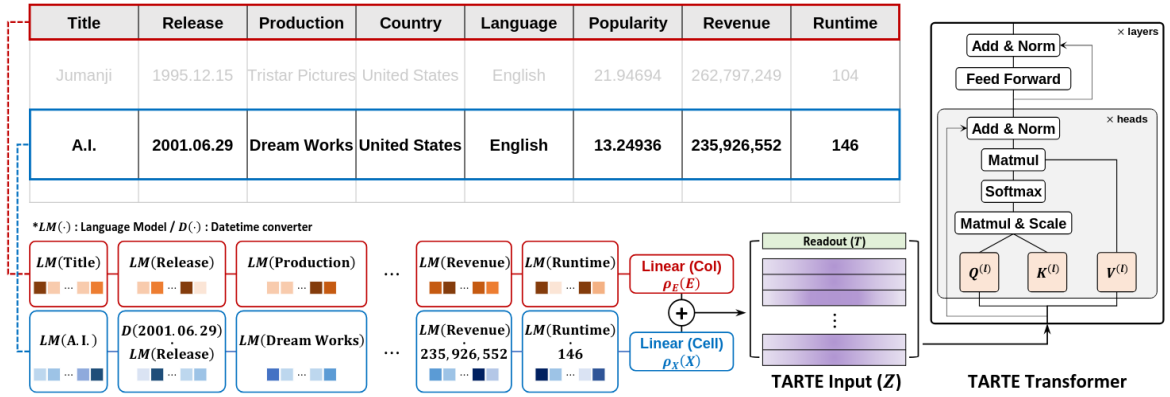
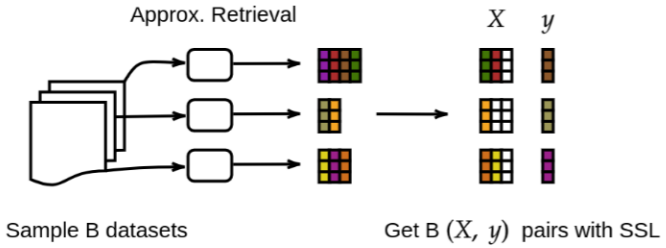


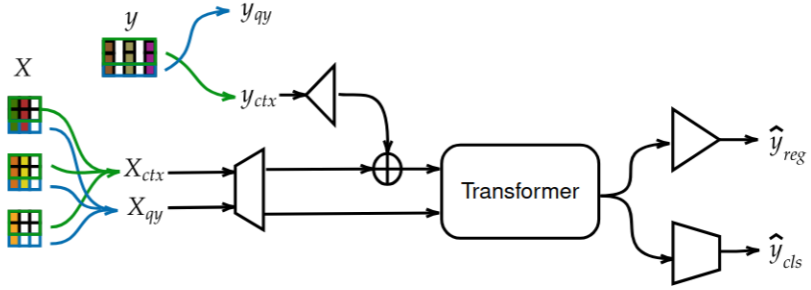
Figure 1: **Transformer-based architecture of TARTE**. TARTE models a row in a table as a set of column(E)–cell(X) pairs. Given a tabular data with multiple data types, TARTE maps the representation of column and cell values to the same dimension using a language model (LM). From the mapped input, the transformer takes in a combination of both column and cell information to contextualize the cell content.

Myung, Jun Kim et al. "Table Foundation Models: on knowledge pre-training for tabular learning", ICML (2025).

TabDPT



(a) Selecting a training batch



(b) Overview of the architecture

layer 6 AI at TD

Core features

- Combines ICL retrieval with self-supervised masked modeling
- Uses real datasets in pretraining

Strengths

- Faster convergence with real-data pretraining
- Strong zero-shot generalization

Limitations

- Less explicit prior modeling than PFN-family
- Pretraining cost higher due to using primarily real data

Ma, Junwei et al. "TabDPT: Scaling Tabular Foundation Models on Real Data", Neurips (2025).

SNT

The Rise of Tabular Foundation Models

A Zoom on TabPFN

A Tour of State-of-the-Art TFMs

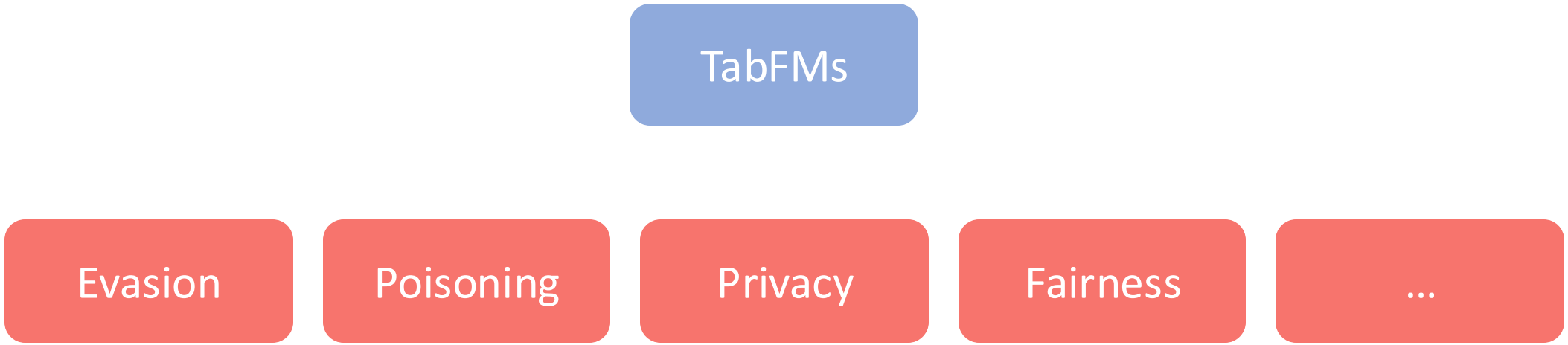
Evasion Attacks on TFMs



UNIVERSITY OF LUXEMBOURG

Trustworthiness of TabFM

As Tabular Foundation Models adoption grows, so do concerns about its trustworthiness



Evasion : Adversarial manipulations of input features at inference time that subtly alter a data record (e.g., credit or medical data) to cause a model to output an incorrect prediction without changing the record's true underlying meaning.

Problem: Evasion Attacks on *Tabular* ML Models

Original

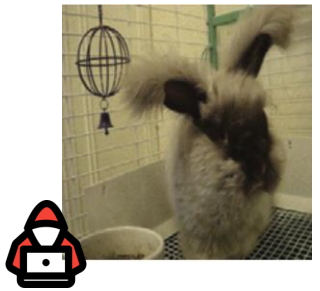


Angora: 91%

\+

Bounded Perturbation

Adversarial



Strainer 77%

Age	Average Heart Rate	Resting Heart Rate	Calories	Active Minutes	Steps
72	135	78	210	45	5200

Constraint Validator

Heart Condition
1

A Li, Yingwei, et al. "Learning transferable adversarial examples via ghost networks." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 07. 2020.

Problem: Evasion Attacks on Tabular ML Models

Manipulating Inputs to fool AI Models

Age	Average Heart Rate	Resting Heart Rate	Calories	Active Minutes	Steps
72	135	78	210	45	5200



Age	Average Heart Rate	Resting Heart Rate	Calories	Active Minutes	Steps
72	110	78	210	50	5200



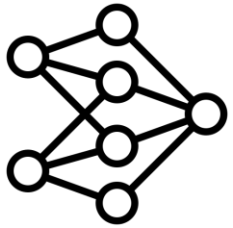


Constraint violated
AHR≈ RHR×[1.5, 2.5]



Age	Average Heart Rate	Resting Heart Rate	Calories	Active Minutes	Steps
72	119	78	210	50	5200





Heart Condition
0

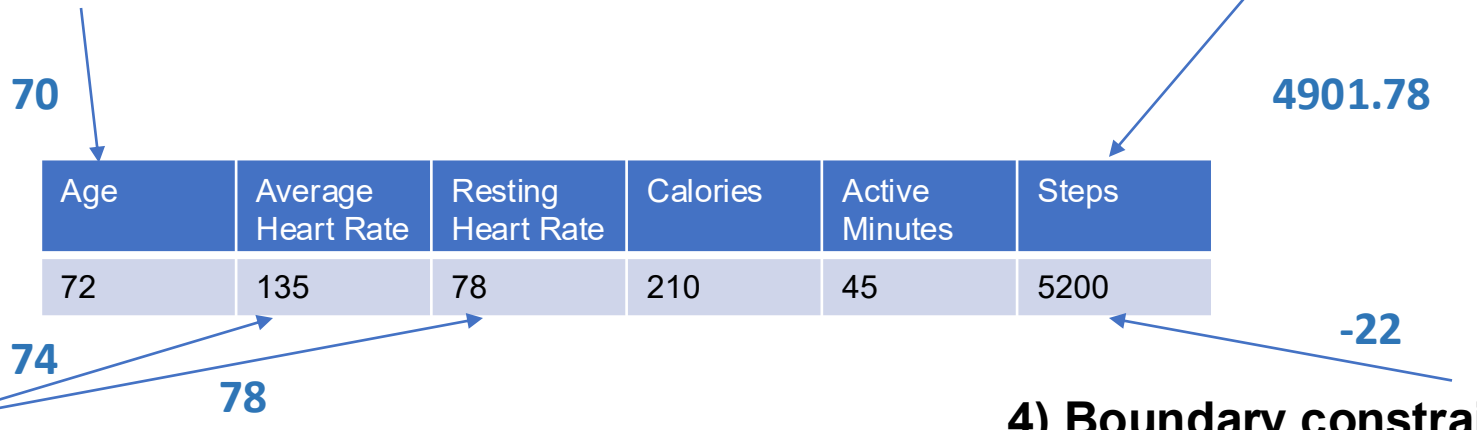
Flip the label
1 -> 0

Problem: Evasion Attacks on Tabular ML Models

Constraint Types:

1) Immutability constraints

2) Type constraints



3) Feature relationship constraints : AHR > RHR

4) Boundary constraints: Upper and lower bounds

Adversarial examples must remain valid and realistic within the domain of the tabular data.

Research Questions

1. How vulnerable are Tabular Foundation Models (TabFMs) such as TabPFNv2 and TabICL to evasion attacks ?

2. How can they be hardened against such threats?

How Vulnerable are TFMs ?
Setup

- *Adversary's Goal:*
  - Flip model predictions (e.g., Rejected → Approved)
  - Stay within perturbation budget L2 ($\varepsilon \leq 0.5$)

- *Adversary's Knowledge (Worst-Case):*
  - White-box access: knows model weights & context
  - Can compute gradients directly

- *Adversary's Capabilities:*
  - *Only modifies test inputs*
  - *Uses SoTA evasion attacks*
  - *Must respect domain constraints*

How Vulnerable are TFMs ?
Setup

Datasets

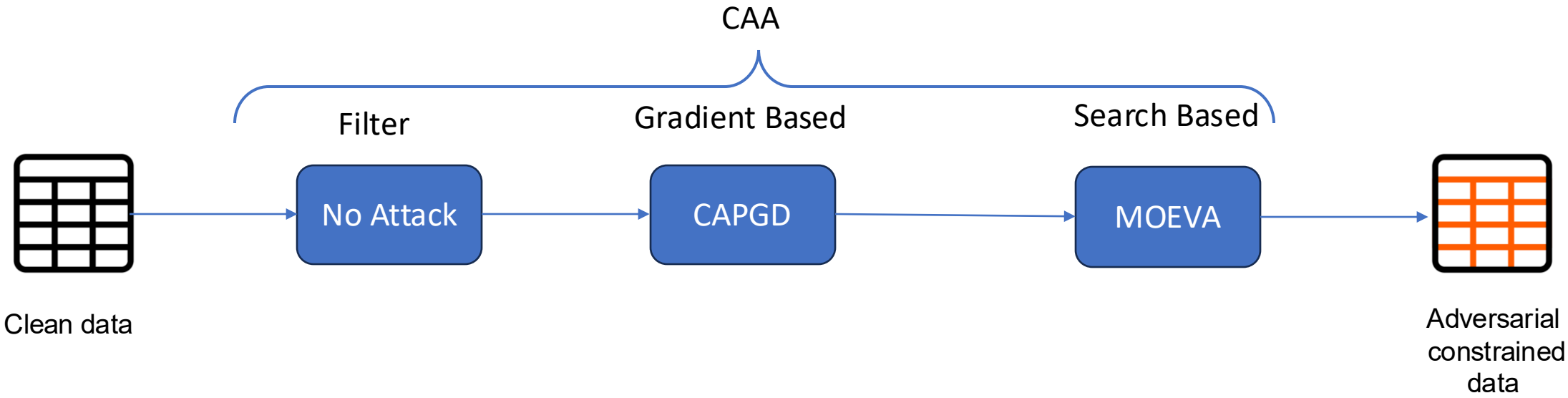
Dataset	Task	Properties Size	# Features	Balance
LCLD [31]	Credit Scoring	1 220 092	28	80/20
URL [32]	Phishing URL detection	11 430	63	50/50
WIDS [30]	ICU patient survival	91 713	186	91.4/8.6

Models

Family	Model
Foundational	TabPFNv2, TabICL
Transformer	TabTransformer
Transformer	TabNet
Regularization	RLN
Regularization	STG
Encoding	VIME
Tree-based	XGB
Tree-based	RF

How Vulnerable are TFMs ?
Setup

Adversarial evasion attack: Constrained Adaptive Attack CAA



Metrics:

We consider the adversarial accuracy of the **positive class of interest, i.e. the recall**

$$\frac{1}{N}\sum_i [y_{1,true_i} = y_{pred_i}]$$

Simonetto, Thibault, Salah Ghamizi, and Maxime Cordy. "Constrained Adaptive Attack: Effective Adversarial Attack Against Deep Neural Networks for Tabular Data." Advances in Neural Information Processing Systems 37 (2024): 27817-27849.

How Vulnerable are TFMs ?
Experimental Results

	Dataset	LCLD			URL			WIDS		
	Model	Clean	CAPGD	CAA	Clean	CAPGD	CAA	Clean	CAPGD	CAA
Classical Models	RLN	68.1 (25.3)	0.2$_{\pm 0.1}$	0.0$_{\pm 0.1}$	94.4 (89.1)	11.5$_{\pm 0.3}$	10.6$_{\pm 0.4}$	77.5 (37.7)	62.1$_{\pm 0.4}$	60.9$_{\pm 0.4}$
	STG	66.0 (24.4)	56.1$_{\pm 0.3}$	52.6$_{\pm 0.3}$	93.3 (83.9)	71.5$_{\pm 0.2}$	56.9$_{\pm 0.3}$	77.6 (36.3)	64.2$_{\pm 0.6}$	63.6$_{\pm 0.7}$
	TabNet	67.5 (26.0)	7.3$_{\pm 0.4}$	0.3$_{\pm 0.2}$	93.4 (89.2)	17.9$_{\pm 0.9}$	9.1$_{\pm 0.4}$	79.7 (36.7)	7.9$_{\pm 0.3}$	5.5$_{\pm 0.5}$
	TabTr.	70.2 (25.2)	36.9$_{\pm 0.9}$	5.6$_{\pm 0.4}$	93.6 (88.0)	9.4$_{\pm 0.2}$	8.3$_{\pm 0.2}$	75.5 (38.5)	47.5$_{\pm 0.6}$	46.0$_{\pm 0.5}$
	VIME	67.1 (24.9)	3.6$_{\pm 0.4}$	2.8$_{\pm 0.4}$	92.5 (85.6)	54.4$_{\pm 0.2}$	47.1$_{\pm 1.3}$	72.3 (38.7)	50.7$_{\pm 0.3}$	50.3$_{\pm 0.3}$
	RF	64.6 (25.6)	N/A	27.5$_{\pm 0.6}$	95.9 (90.3)	N/A	48.3$_{\pm 1.6}$	49.3 (43.8)	N/A	11.0$_{\pm 0.4}$
	XGB	67.9 (26.1)	N/A	10.3$_{\pm 0.7}$	97.4 (93.8)	N/A	16.3$_{\pm 0.9}$	80.4 (40.5)	N/A	35.7$_{\pm 0.7}$
TabFMs	TabPFNv2	59.7 (25.4)	22.4 $_{\pm 0.1}$	9.0 $_{\pm 0.2}$	97.1 (92.8)	47.6 $_{\pm 0.2}$	25.2 $_{\pm 0.5}$	70.7 (43.0)	49.2 $_{\pm 0.4}$	30.8 $_{\pm 0.4}$
	TabICL	59.6 (25.5)	18.5 $_{\pm 0.4}$	10.2 $_{\pm 0.1}$	96.7 (92.8)	82.9 $_{\pm 0.2}$	20.2 $_{\pm 0.3}$	75.5 (42.0)	38.2 $_{\pm 0.3}$	30.5 $_{\pm 0.2}$

Recall against Clean data, CAPGD and CAA adversarial examples. For Clean data we also give Mathew Correlation Coefficient (MCC) in parenthesis. A **lower** robust performance means a more **effective** attack.

Answer to RQ1:

Our evidence shows that Tabular FMs are *at least* vulnerable to evasion attacks as their tabular non-FMs models counterparts.

RQ2: Hardening of TabFMs: Problem Formulation

AFT Adversarial FineTuning:

$$\min_{\theta} \mathbb{E}_{(X_{test},Y_{test})\sim D} [\max_{\delta\in\Delta} \mathcal{L}(h(X_{test}+\delta|X_{train},Y_{train},\theta),Y_{test})]$$

Adversarial perturbation

Weights of models

Loss Function

Budget

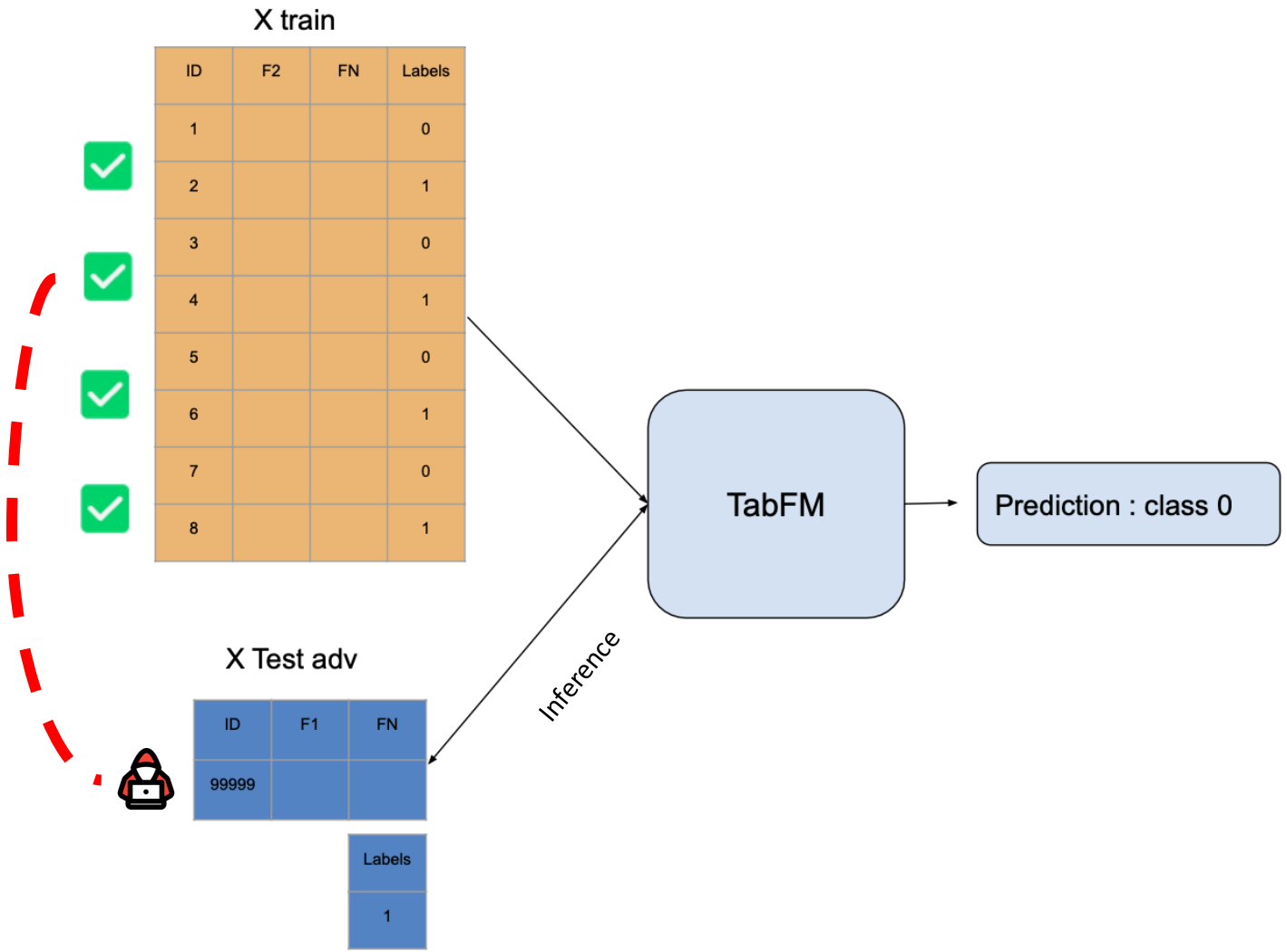
(Ours) AICL Adversarial In-Context-Learning:

$$\min_{\delta^*\in\Delta^*} \mathbb{E}_{(X_{test},Y_{test})\sim D} [\max_{\delta\in\Delta} \mathcal{L}(h(X_{test}+\delta|X_{train}+\delta^*,Y_{train},\theta),Y_{test})]$$

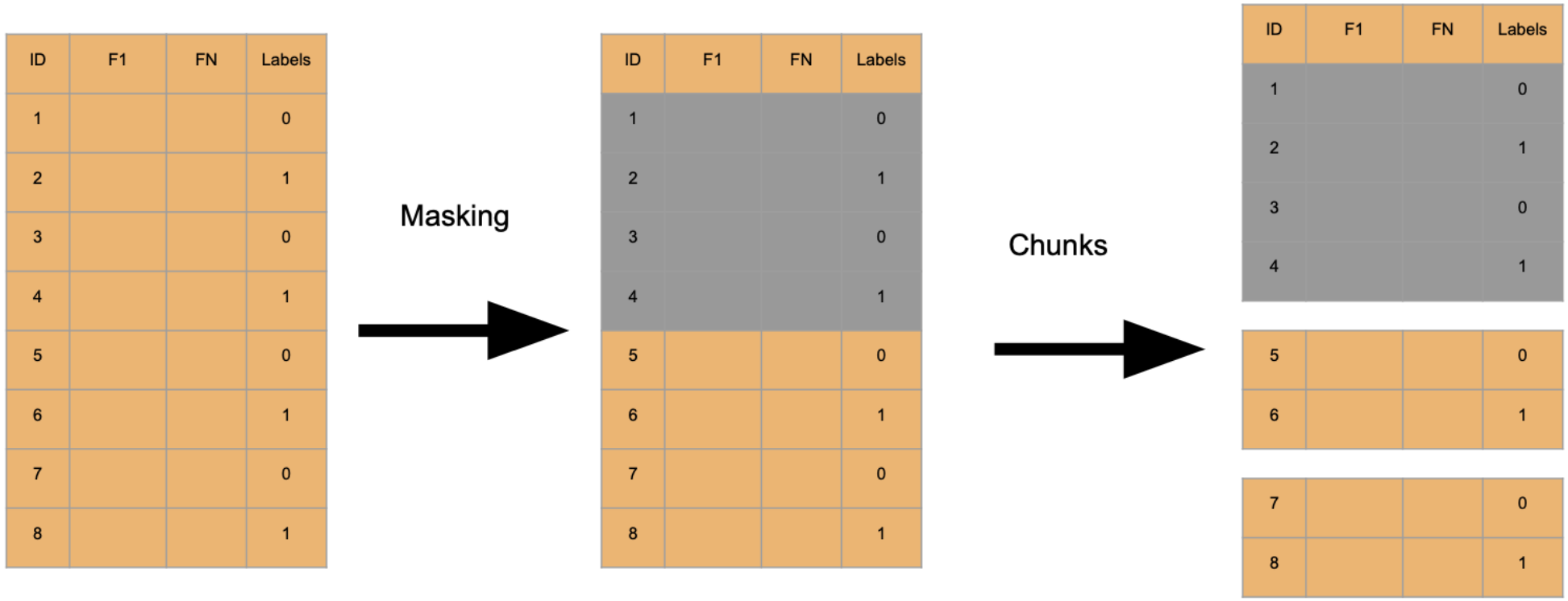
Hardening of TabFMs:
A Distribution Mismatch ?

TabFM hypothesis breakdown:

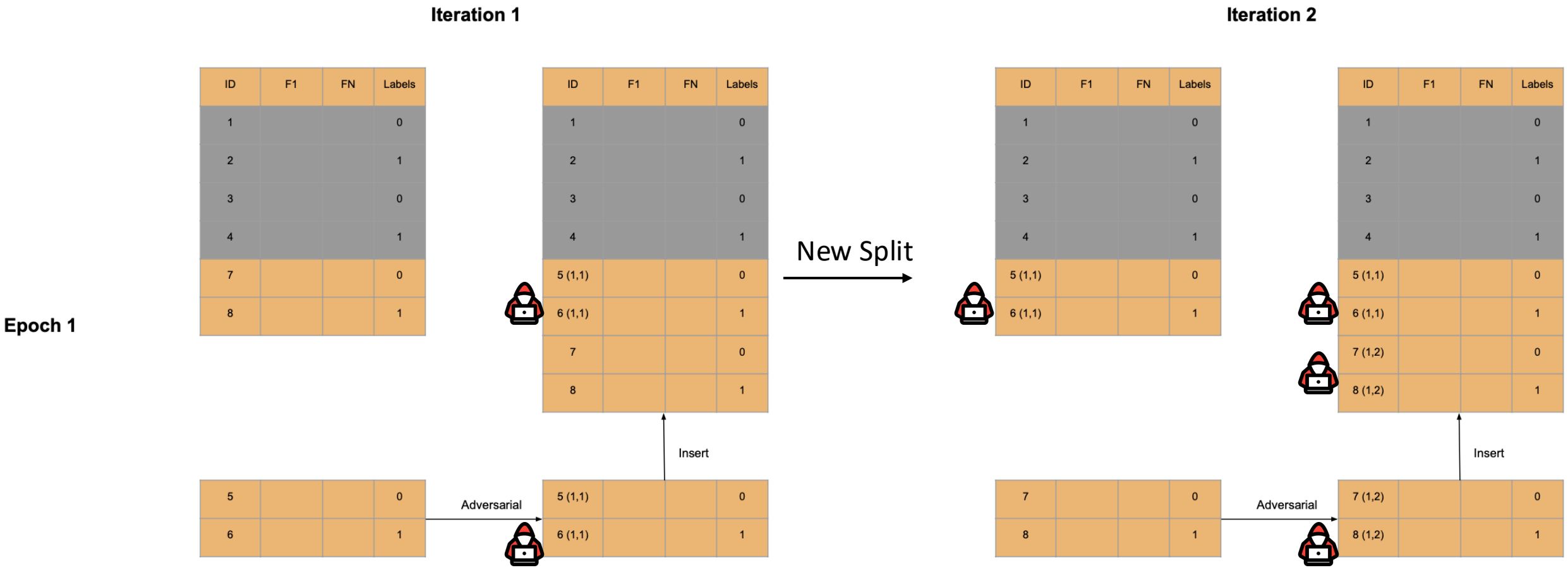
$$Dist(X_{ctx}) \neq Dist(X_{qy})$$



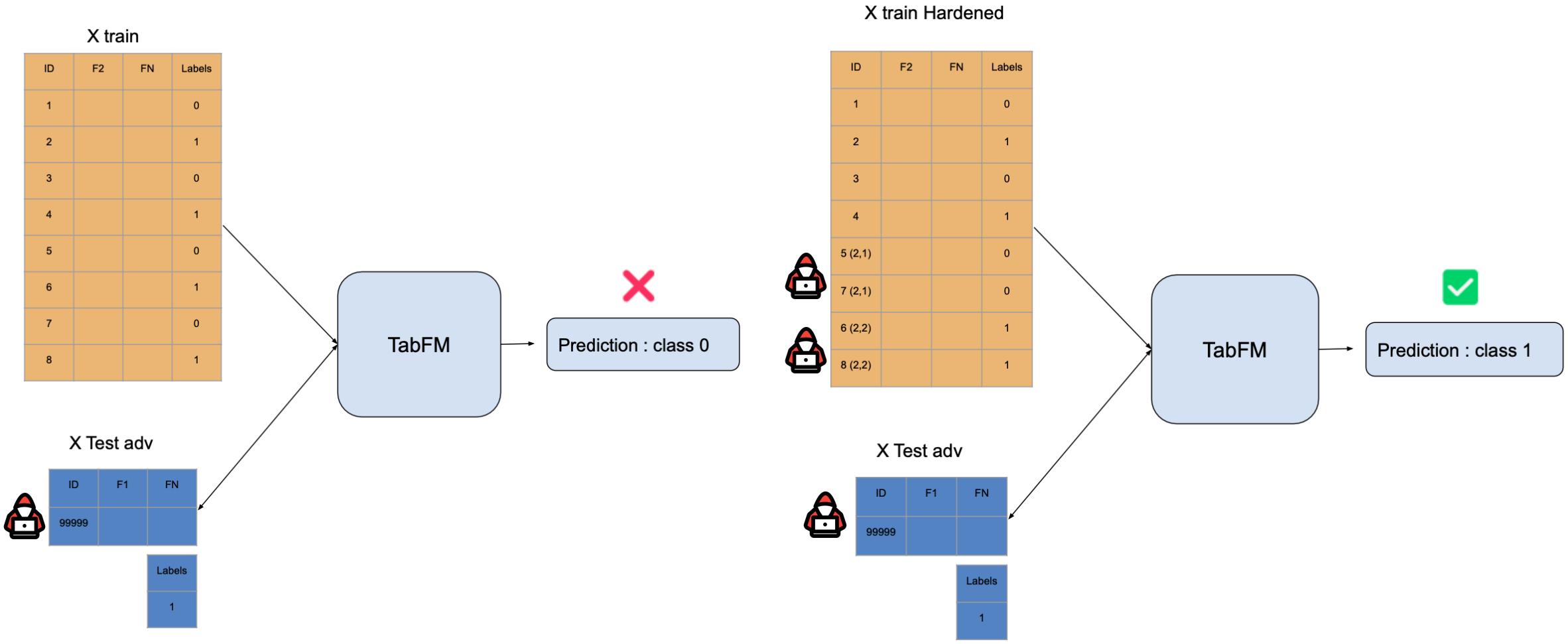
Hardening of TabFMs: the AICL Method (1/3)



Hardening of TabFMs: the AICL Method (2/3)



Hardening of TabFMs: The AICL Method (3/3)



Hardening of TabFMs: Results

Attacks used in hardening

Attacks used during evaluation

Dataset		LCLD			URL			WIDS		
Model	Training	Clean	CAPGD	CAA	Clean	CAPGD	CAA	Clean	CAPGD	CAA
TabPFN2	Original	59.7	22.4 ±0.1	9.0 ±0.2	97.1	47.6 ±0.2	25.2 ±0.5	70.7	49.2 ±0.4	30.8 ±0.4
	AICL CAA	60.4	49.3 ±0.3	41.2 ±0.0	96.2	96.1 ±0.0	69.4 ±0.2	70.7	65.7 ±0.3	51.5 ±0.2
	AICL CAPGD	55.8	42.0 ±0.5	14.7 ±0.2	95.6	94.7 ±0.1	25.9 ±0.2	70.7	63.7 ±0.2	42.5 ±0.4
	AFT CAA	62.7	20.6 ±0.4	12.0 ±0.1	97.2	51.1 ±0.2	31.2 ±0.3	73.3	38.2 ±0.3	31.1 ±0.5
	AFT CPGD	61.6	22.3 ±0.3	12.8 ±0.1	97.1	53.2 ±0.4	31.8 ±0.1	68.7	34.2 ±0.3	25.7 ±0.2
TabICL	Original	59.6	18.5 ±0.4	10.2 ±0.1	96.7	82.9 ±0.2	20.2 ±0.3	75.5	38.2 ±0.3	30.5 ±0.3
	AICL CAA	60.3	51.2 ±0.3	39.6 ±0.2	97.3	97.0 ±0.0	75.6 ±0.2	76.2	60.0 ±0.2	54.7 ±0.2
	AICL CAPGD	58.8	47.1 ±0.2	20.8 ±0.2	96.4	95.8 ±0.1	31.0 ±0.4	75.5	55.5 ±0.2	50.3 ±0.2
	AFT CAA	54.4	22.6 ±0.1	20.4 ±0.2	96.6	80.3 ±0.2	22.6 ±0.2	71.6	44.1 ±0.3	32.2 ±0.4
	AFT CPGD	59.2	18.4 ±0.2	4.3 ±0.1	96.7	84.8 ±0.3	16.8 ±0.1	54.6	35.2 ±0.4	14.8 ±0.3

The Higher the Better

Robust accuracy on Clean data and adversarial CAPGD/CAA attacks using AICL and AFT hardening methods

Answer to RQ2:

Robustness can be increased via the context only, by carefully replacing clean samples with their adversarial counterparts



Conclusion & QA

Conclusion

- Tabular Foundation Models are now competitive, especially in small-data regimes.
- The shift is from per-dataset training to cross-dataset pretraining.
- Pretraining data design (synthetic vs real) is now central to performance.
- The field is rapidly diversifying in architectures and training strategies.
- Robustness remains unresolved: TFMs are vulnerable and require dedicated hardening.

Q&A

Interdisciplinary Centre for Security, Reliability and Trust

Contact:



Karim Tit

Postdoctoral Researcher (SnT)

karim.tit@uni.lu

Connect with us

@SnT_uni_lu

SnT, Interdisciplinary Centre for Security, Reliability and Trust