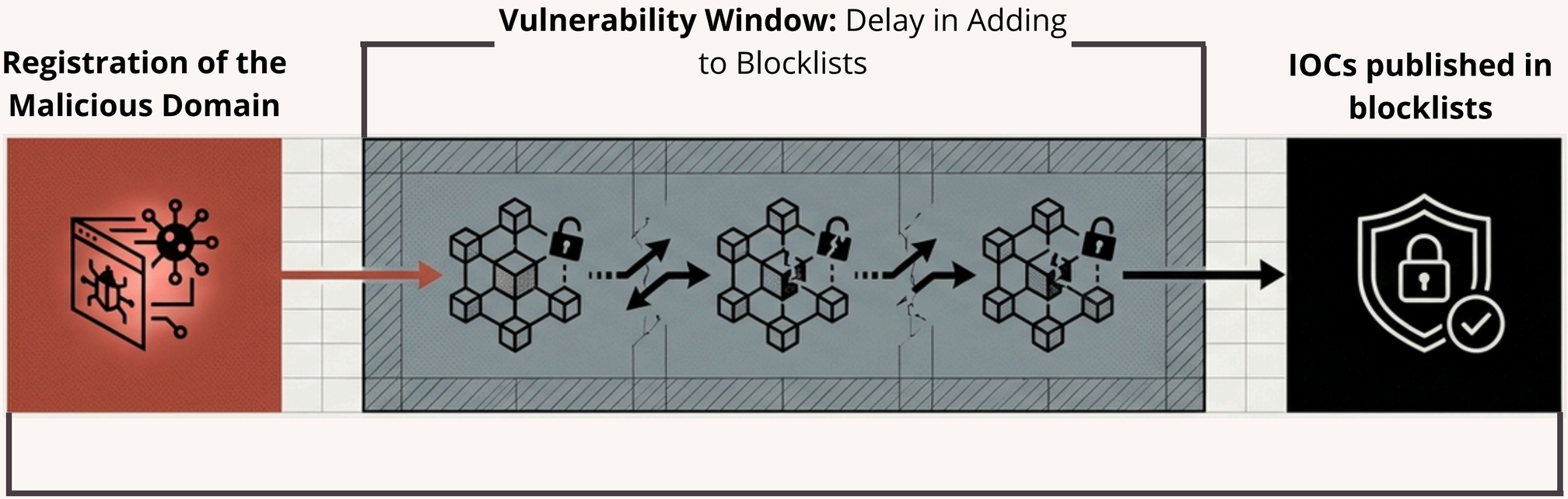
# Malicious Domain Detection in Real-World Academic DNS Traffic: A Machine-Learning Approach

DNS-OARC 47 - 2026

Guilherme Romanholo Bofo
Prof. Dr. Adriano Mauro Cansian

# Problem Context

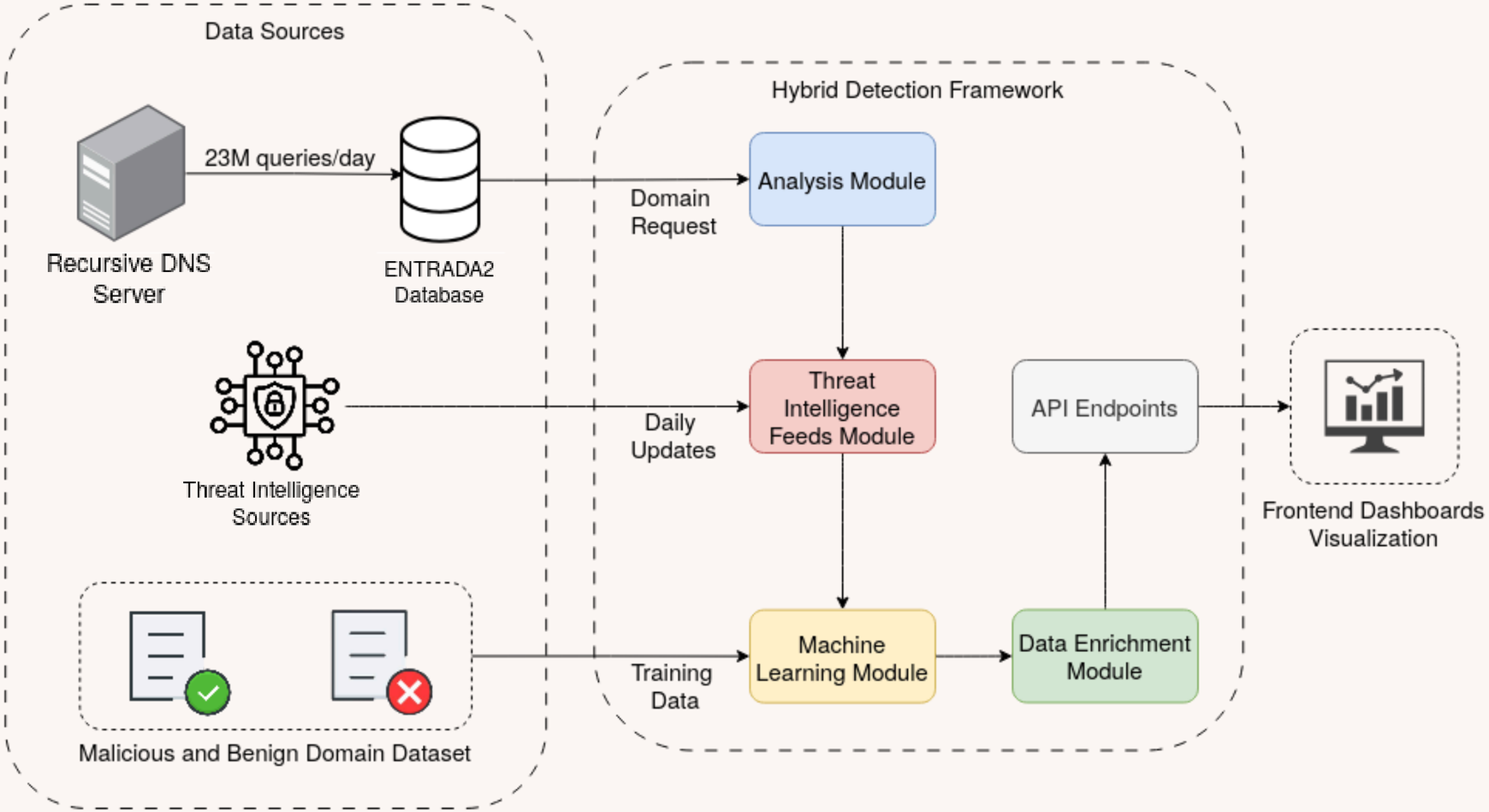



**Timeline of a Malicious Domain**

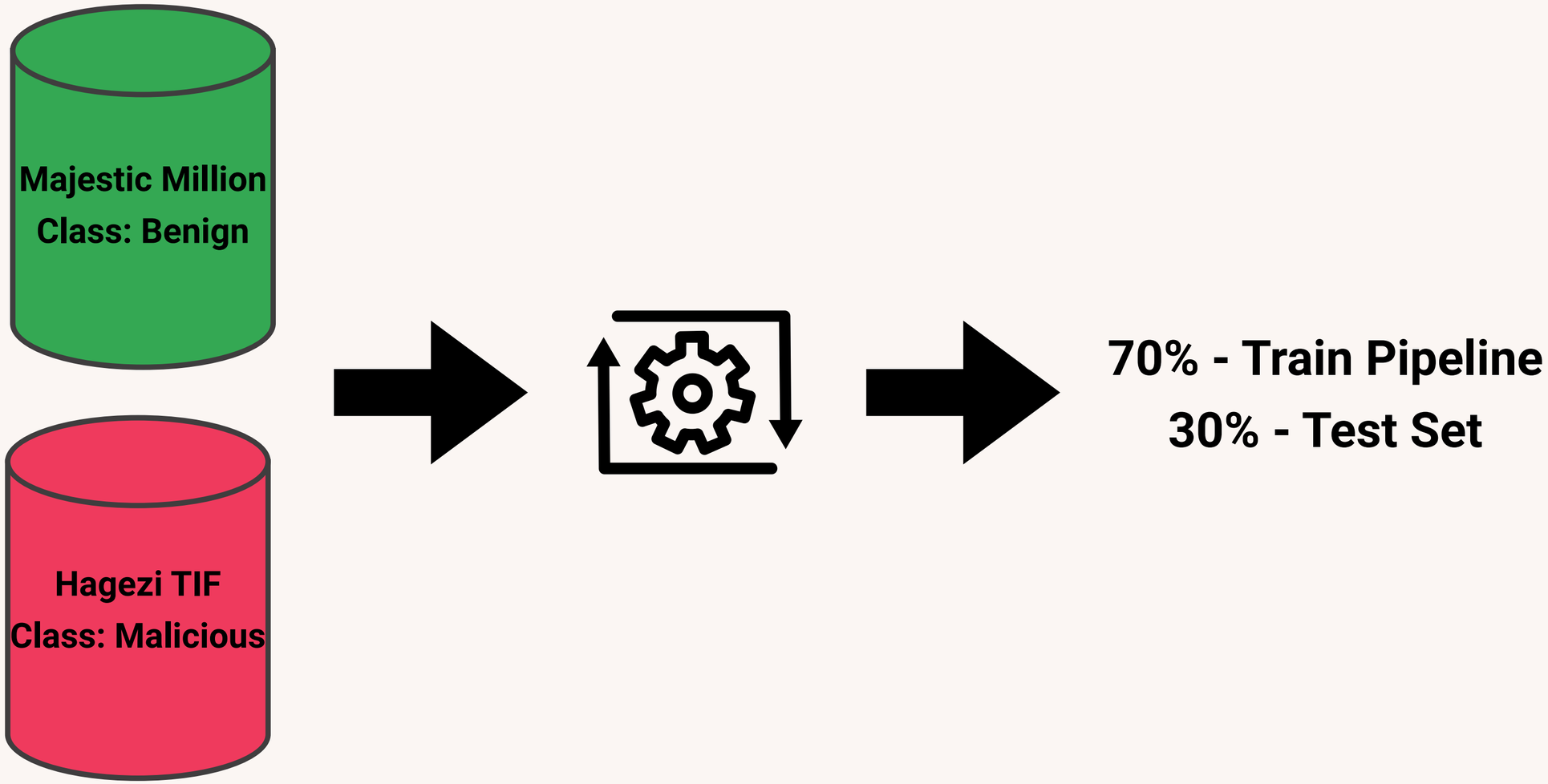
# Objectives

- Verify malicious domain detection with a hybrid combination of threat intelligence feeds and a lightweight Random Forest.
- Evaluate the model using lexical, active DNS and TF-IDF features.
- Analyze system behavior in highly imbalanced real-world traffic, especially regarding early detection and false-positive trade-offs.

# Hybrid Detection Architecuture



# Train Data


Majestic Million
Class: Benign
Hagezi TIF
Class: Malicious
70% - Train Pipeline
30% - Test Set

# Feature Types

## Lexical

Features extracted directly from the syntax and character composition of domain names.

## Active DNS

Live operational metrics collected dynamically through Cloudflare DoH (DNS Over HTTPS).

## TF-IDF

Vectorization metrics specifically mapping textual domain patterns to identify generated anomalies.

# Test-Set Results

| Model State | Precision [1] | Recall [2] | F1 Score [3] | AUC [4] |
|---|---|---|---|---|
| B (Baseline) | 0.7594 | 0.8083 | 0.7831 | 0.8877 |
| F (Fine-tuned) | 0.7688 | 0.8309 | 0.7986 | 0.9028 |
| BT (Base + TF-IDF) | 0.7897 | 0.8211 | 0.8051 | 0.9070 |
| FT (Fine-tuned + TF-IDF) ✓ | 0.7940 | 0.8318 | 0.8125 | 0.9176 |

[1] Measures the proportion of detected malicious domains that are truly malicious

[2] Measures the proportion of actual malicious domains correctly identified by the model

[3] The harmonic mean of precision and recall, providing a balance between both metrics

[4] Measures the model's ability to distinguish between malicious and benign domains across different classification thresholds

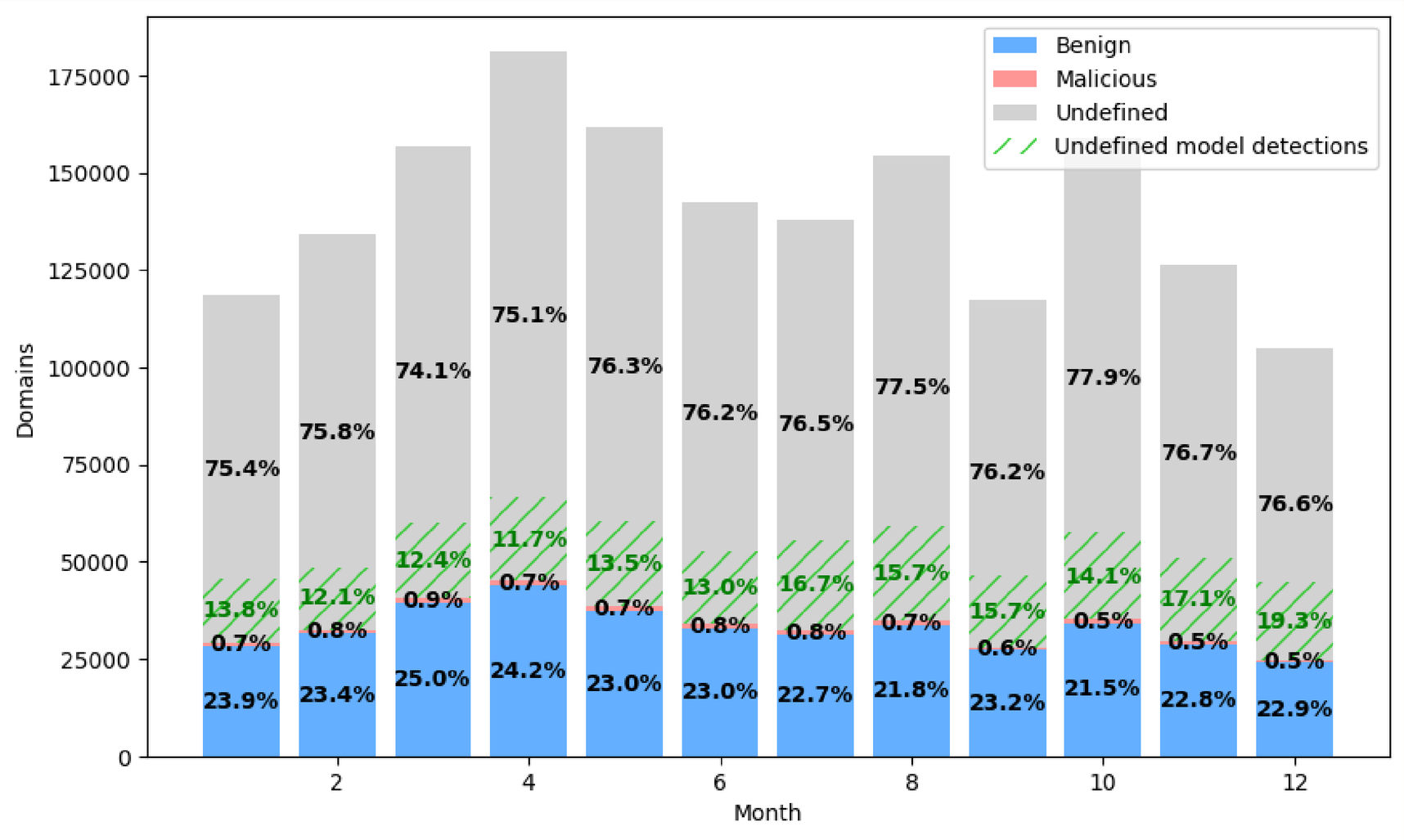
# Real-World Results (1)

- The proposed system was evaluated in a real academic recursive resolver (UNESP) that processes **more than 20 million DNS queries per day**.

- During the one-year deployment, the model identified **between 512 and 1439 malicious domains per month**, all of which were confirmed to be present on public blocklists

- In operational conditions, **precision** ranged from **0.142 to 0.364**, and **recall** ranged from **0.434 to 0.667**.

# Real-World Results (2)



# Threats to Validity

- **Internal validity**: labels depend on external threat intelligence feeds, which may contain noise or delays.
- **External validity**: evaluation was based on a single academic recursive resolver and may differ in corporate environments.
- **Construct validity**: most observed domains were undefined, which may bias the quantitative assessment.
- **Operational validity**: low precision limits the use of the system for fully automatic blocking without human oversight.

# Conclusions

- The hybrid combination is operationally feasible for passive DNS monitoring in a production academic network.
- The tuned model with TF-IDF obtained the best offline performance, reaching AUC = 0.918 and F1 = 0.8125.
- The proposed system should be interpreted as an early-warning and triage component that complements blocklists rather than replaces them.

# Future Work

- Reduce false positives through threshold calibration, ensemble refinement and post-processing rules (fuzzy).
- Explore semi-supervised or unsupervised learning to leverage undefined domains.
- Investigate more robust architectures while preserving real-time feasibility.
- Develop specialized models for phishing, malware and DGA domains instead of a single general-purpose classifier.

# Acknowledgements

This work was formally published and presented at the 28th International Conference on Enterprise Information Systems (ICEIS).

And this work was conducted with the support of the Brazilian Network Information Center (NIC.br) – FUNDUNESP Process No. 3467/2023.

unesp

ACME!
Cybersecurity Research

# Thank you!

Any questions?

guilherme@acmesecurity.org