

# Explainable Malware Analysis

A Report from RRP EMA Project



**[RECOVERY  
AND RESILIENCE]  
PLAN**

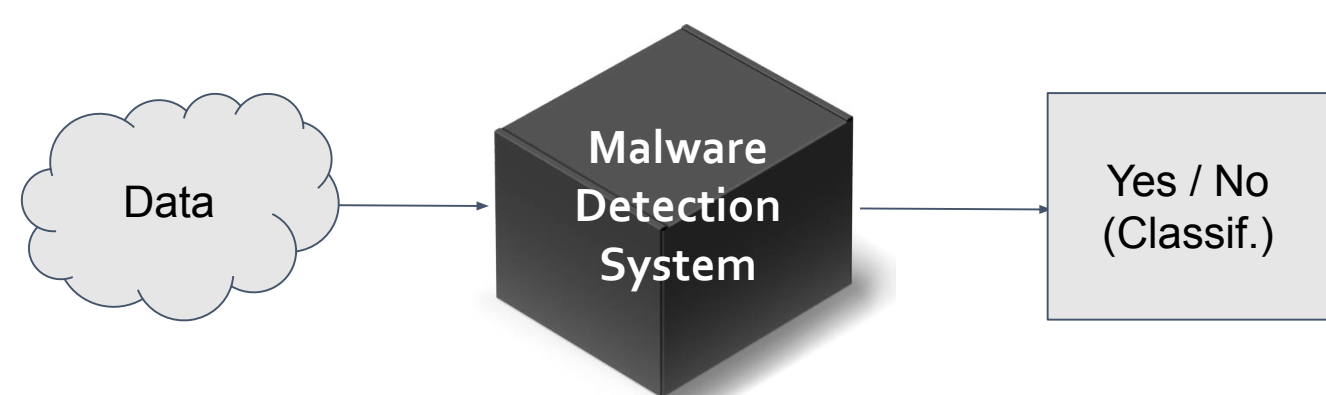
project no. 09105-03-V02-00064

Growing Volumes of  
Malware

## Too much data:

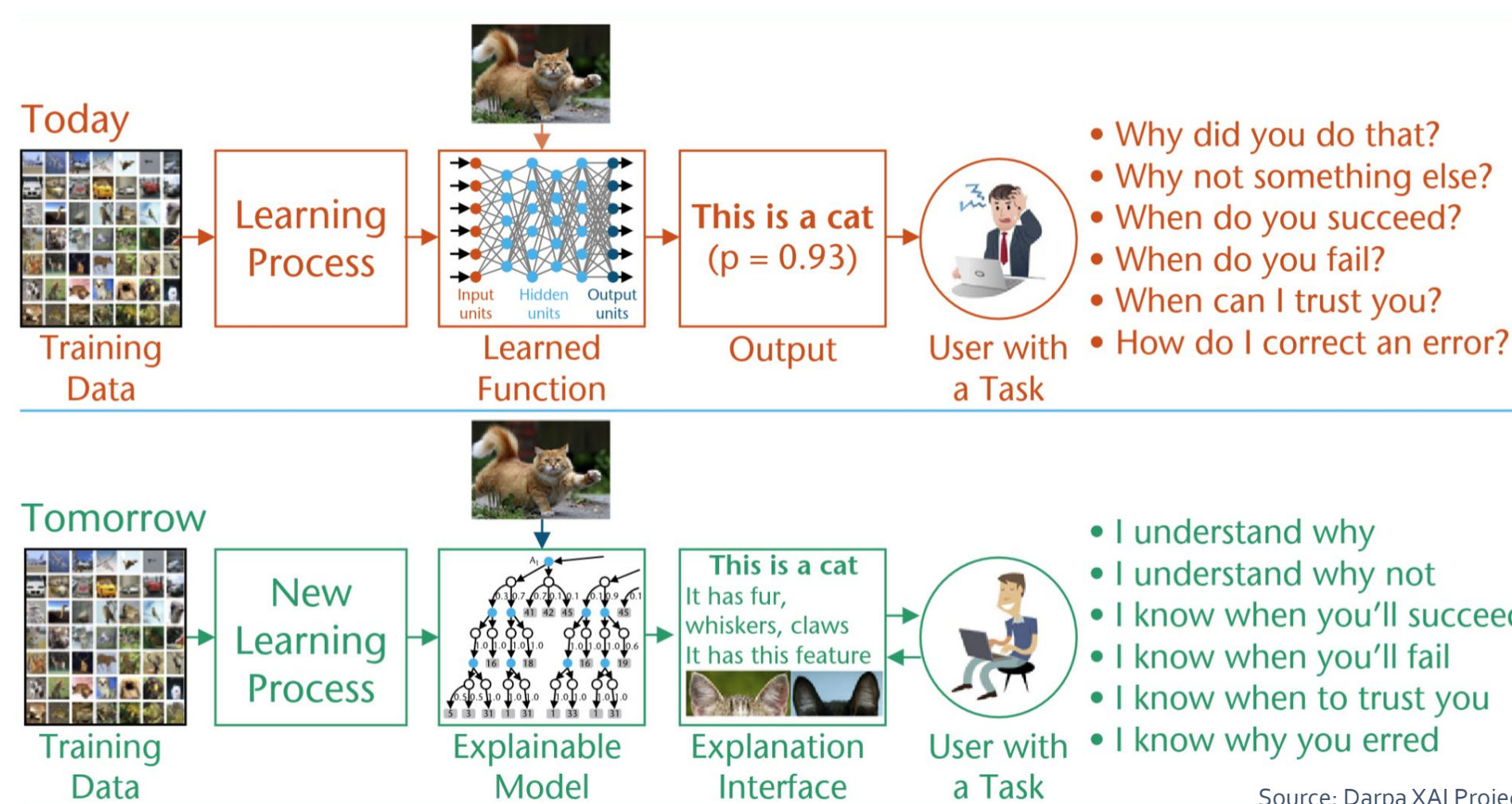
- 60+ million new malware strains
- 300,000+ daily

(AVG.com, 2024)



Black-box ML established in malware analysis

Black-Box vs.  
eXplainable AI



## Post-hoc methods: LIME and SHAP

- Applied on top of any black-box ML
- Limited insights (feature importance)
- Fidelity issues (only approximate)

XAI Methods

## Symbolic Methods: Decision-tree learning, Concept Learning

- Good results but slow
- Explanation is a logical formula (human-readable, directly acts as symbolic classifier = fidelity 100%)

## Neuro-Symbolic Methods: Knowledge-base embedding, logic explained networks, concept recognition...

- Similar but more efficient

Dynamic Malware  
Analysis

**Datasets:** Suitable dataset available (EMBER 3,2 million samples, SOREL 15 million samples) but only for static malware analysis

## Static Malware Analysis:

- Does not require to run code
- Easier but fewer details and less efficient

## Dynamic Malware Analysis:

- Analyses the behaviour of actual running code
- More details may help XAI methods to compete with black-box methods

Data Representation and  
Organization

## Malware Ontology:

- Semantic data model
- Common language to express explanations

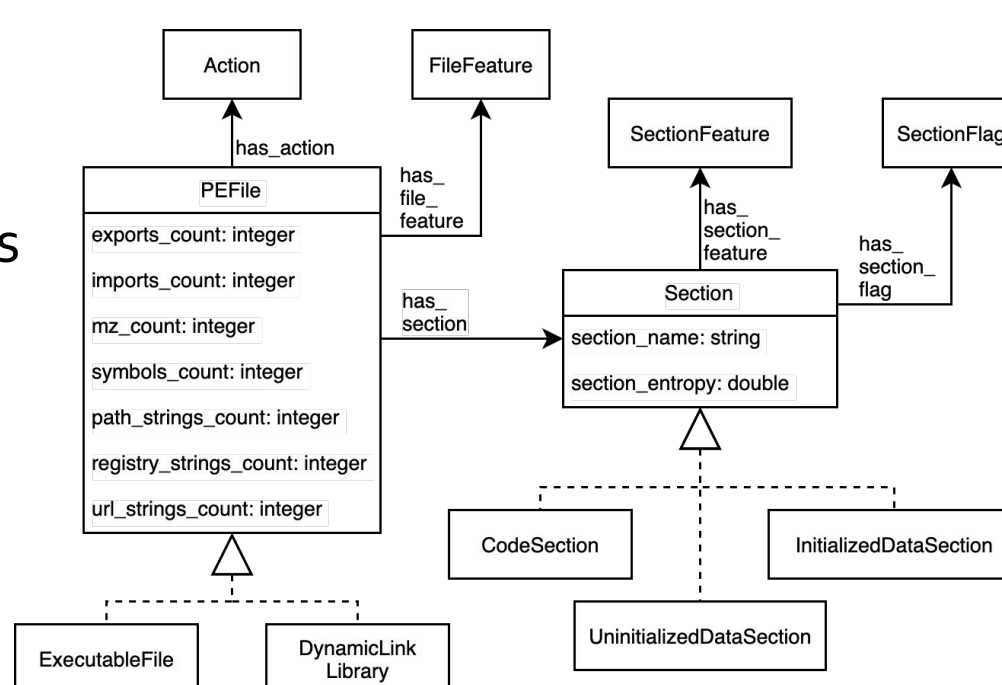
## Data break-down:

### Organic:

- Based on Malware families

### By clustering:

- Recognize similar structures in data
- Smaller datasets with compressed information



Explanations and User  
Studies

## Large palette of learned explanations:

- Shorter vs longer formulas
- Different languages, different expressivity
- How to present to human users?

$ExecutableFile \sqcap \exists has\_section. \exists has\_section\_feature. (HighEntropy \sqcup WriteExecuteSection)$

"The instance is an executable (EXE) file and it has a section either with high entropy or that is jointly writable and executable."



**User studies:** Understand what types and expressivity of explanations is the most useful to malware analysts



STU KINiT



FAKULTA MATEMATIKY,  
FYZIKY A INFORMATIKY  
Univerzita Komenského  
v Bratislave

MATEFYZ  
CONNECTIONS

Martin Homola<sup>1</sup>, Štefan Balogh<sup>2</sup>, Ján Klůka<sup>1</sup>, Daniela Chudá<sup>3</sup>, Iveta Bečková<sup>1</sup>, Jaroslav Kopčan<sup>3</sup>  
(and collaborators)

<sup>1</sup> Comenius University in Bratislava

<sup>2</sup> Slovak University of Technology in Bratislava

<sup>3</sup> Kempelen Institute of Intelligent Technologies