



**AI-Driven Cybersecurity for  
Financial Service Providers**

## **Module 2 EMTD in Depth: The Five-Stage Detection Pipeline**

CyberAid Learning Centre – EMTD Tutorial Course

October 2026

# EMTD at a glance

5

pipeline stages

24

attack patterns in the  
ontology

12

always-run deterministic  
checks

69

detection components

8

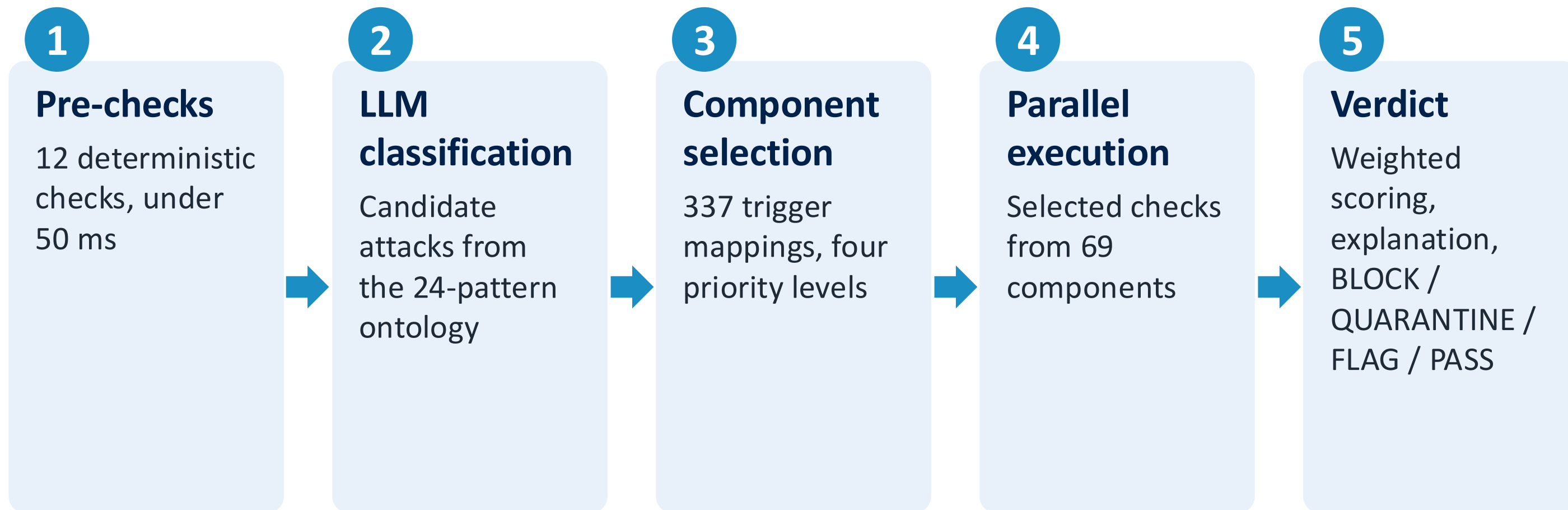
functional families

337

attack-to-component  
trigger mappings

*Deterministic rules + LLM analysis, fused into a BLOCK / QUARANTINE / FLAG / PASS verdict.*

# The five-stage pipeline



# Stage 1: Deterministic pre-checks

**< 50**

**ms**

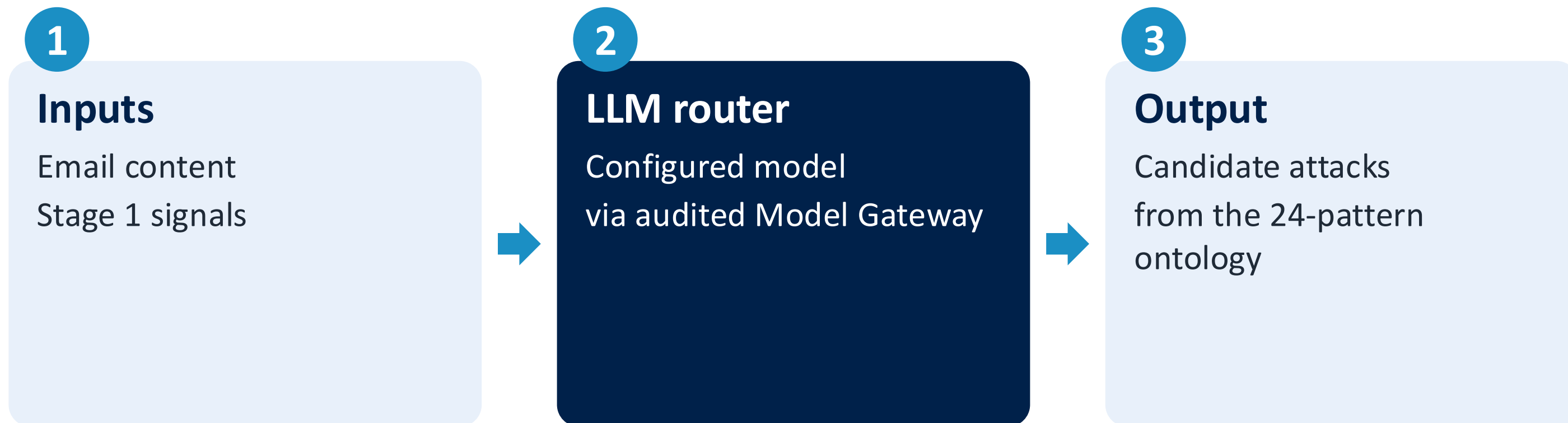
for the 12 always-run  
checks

## Conditional checks

URL, calendar, IBAN and  
regulatory checks run when  
relevant

- SPF, DKIM and DMARC results, as supplied
- Reply-To mismatch and subject anomalies
- Unicode homoglyphs and zero-width characters
- Display-name consistency and domain similarity
- Known contacts, first-time contacts, sender email changes

# Stage 2: LLM attack classification



*Candidates are proposals that drive component selection; they are not the final verdict.*

# Stage 3: Dynamic component selection

337

attack-to-component  
trigger mappings

CRITICAL

HIGH

MEDIUM

LOW

*Four routing priority levels*

- Select components linked to the candidate attacks
- Deduplicate the selection
- Always retain the 12 always-run checks
- Avoid running all 69 components on every email

# Stage 4: Parallel specialist components

| Family | Focus                                     | Components |
|--------|-------------------------------------------|------------|
| SA-01  | Header authentication and transport       | 9          |
| SA-02  | Content, semantics and social engineering | 12         |
| SA-03  | Sender identity and contact consistency   | 8          |
| SA-04  | Behavioural anomalies                     | 7          |
| SA-05  | Files and attachments                     | 9          |
| SA-06  | URLs and links                            | 8          |
| SA-07  | Financial and regulatory processes        | 9          |
| SA-08  | OSINT and organisational context          | 7          |

*Methods: deterministic, heuristic or model-assisted. LLM analysis is an execution method, not a family.*

# Stage 5: Orchestration and verdict

| PASS<br>0.00 – < 0.35                               | FLAG<br>0.35 – < 0.60                | QUARANTINE<br>0.60 – < 0.85     | BLOCK<br>0.85 – 1.00               |
|-----------------------------------------------------|--------------------------------------|---------------------------------|------------------------------------|
| Below review threshold; check analysis completeness | Suspicious indicators warrant review | Hold message for analyst review | Highest threat: block and escalate |

- Weighted confidence scoring is deterministic
- Analyst explanation is model-assisted, with a deterministic fallback
- Missing inputs, unavailable checks and degraded analysis are reported
- A verdict expresses an assessment; delivery follows the gateway policy



# Scoring I: weighted confidence fusion

**4**

severity weight:  
CRITICAL

**3**

severity weight:  
HIGH

**2**

severity weight:  
MEDIUM

**1**

severity weight:  
LOW

**Base score =  $\Sigma$  (confidence  $\times$  severity weight) of suspicious findings  
 $\div$   $\Sigma$  severity weight of all valid findings, including clean ones**

Skipped, stubbed and erroneous results are excluded. Corroboration, capping, clean-signal discounts and minimum-score rules then produce the final score (0.00 – 1.00).

# Scoring II: safeguards against noise

## Single-family cap

All suspicious findings in one family:  
score capped at 0.59  
(below the 0.60  
QUARANTINE threshold)

## Corroboration boost

Suspicious CRITICAL  
findings in 2 families:  $\times 1.15$   
In 3 or more families:  $\times 1.25$

## Discounts and minimums

Clean-signal discounts  
reduce the score  
Minimum-score rules  
apply  
Skipped or errored  
results excluded

*A noisy signal in one area cannot alone cause a QUARANTINE or BLOCK assessment.*



# Reading the scores honestly

**Rule confidences are configured outputs, not measured accuracy**

**Severity weights are fixed, not learned from history**

**Final score is not a calibrated probability**

**Corroboration is a heuristic; families are not independent evidence**

**Ontology coverage does not establish detection accuracy**

**LLM confidence depends on the actual analysis**



# Key takeaways

- Hybrid design: deterministic rules plus LLM analysis
- Catalogue-driven selection avoids running everything on every email
- Eight component families provide diverse evidence
- Deterministic fusion with corroboration and capping safeguards
- Scores are transparent assessments, not calibrated probabilities



# Thank you!

Connect With CyberAID:  
**LinkedIn: CyberAID**

**Next: Module 3 – Worked Example: Catching a CEO Impersonation (BEC)**



CyberAID has received funding from the DIGITAL EUROPE Programme under grant agreement No. 101249596