



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

Module 4 Positioning EMTD Against the State of the Art

CyberAid Learning Centre – EMTD Tutorial Course

October 2026

The email threat detection landscape

1 Gateways and rule/reputation filters

Authentication checks, blocklists, reputation, signatures, sandboxing. E.g. Microsoft Defender for Office 365, Proofpoint, Mimecast

2 API-based and behavioural email security

Learn normal communication patterns and flag deviations. Integrate with the mail platform

3 ML classifiers in the literature

Feature-based, NLP and deep models for phishing and BEC detection

4 LLM-based and agentic detection

LLMs reason about content and context; emerging academic and industrial work

Typical strengths and limitations

Approach	Typical strengths	Typical limitations
Gateways, rules, reputation	Fast, predictable, well understood	Weak on authentication-clean social engineering
Behavioural / API-based	Detects deviation from normal patterns	Depends on baseline; often opaque reasoning
ML classifiers	Generalise across many messages	Hard to explain; domain adaptation needed
Pure LLM approaches	Understand content and context	Consistency, auditability, cost, data governance

General tendencies, not statements about any specific product.

How EMTD compares on dimensions that matter

Dimension	Gateways / rules	ML classifiers	Pure LLM	EMTD
Detection logic	Rules, reputation	Learned models	Model reasoning	Hybrid: rules + LLM
Explainability	Rule hits	Often limited	Free-text	Scored findings + explanation
Financial semantics	Generic	Domain-dependent	Prompt-dependent	24-pattern ontology, SA-07
Governance	Product-specific	Varies	Varies	Audited Model Gateway
Missing input	Varies	Varies	Varies	Degradation reported

Indicative and qualitative; not a benchmark.



EMTD's distinguishing design choices

Hybrid by design

Deterministic checks for speed and certainty; LLM analysis for semantics

Ontology-driven

24 financial attack patterns route 337 trigger mappings to 69 components

Corroboration-based scoring

Cross-family evidence boosts; single-family evidence is capped

Transparent and auditable

Component findings, analyst explanation, reported degradation, audited Model Gateway

Limitations we state openly

- 1 Score calibration against real data
- 2 Benchmarking on public and financial-sector datasets
- 3 Prompt injection and adversarial content aimed at the LLM stages
- 4 Ontology and trigger-mapping maintenance as attackers evolve
- 5 Latency and cost of model-assisted stages

How to evaluate EMTD

Detection quality

Precision and recall per pattern
Quality per verdict band

Operational cost

False-positive burden on analysts
Time to verdict

Explanation quality

Accuracy of the explanation
Actionability for analysts

Robustness

Adversarial content
Missing inputs and degraded components

Proposed evaluation approach; to be run on realistic financial-sector data and in the CyberAid pilots.

Key takeaways

- Four families of approaches: gateways, behavioural, ML classifiers, LLM-based
- EMTD is hybrid, ontology-driven, corroboration-scored and auditable
- Tailored to wealth-management attack patterns and financial processes
- Calibration, benchmarking and adversarial robustness remain open
- Evaluation should combine detection, cost, explanation and robustness



Thank you!

Connect With CyberAID:
LinkedIn: CyberAID

Next: Module 5 – Integrating EMTD into CyberAid



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