Author: G.Moresi

Date: 09.10.2025

Automated Resilience Index (ARI): A New Metric for Measuring Phishing Mitigation Efficiency using AI

Executive Summary

Traditional phishing awareness programs rely on Click-Through Rate (CTR) metrics to evaluate user performance during phishing simulations. **However, CTR-based metrics have become increasingly meaningless in modern enterprise environments dominated by automated filters, security gateways, and user fatigue.**

This whitepaper introduces a new, mathematically grounded concept — the Automated Resilience Index (ARI) which measures the systemic and automated capacity to neutralize phishing threats before user interaction becomes possible.

The ARI formula captures the true operational performance of cybersecurity automation and response, reflecting speed, automation efficiency, and containment success. It is a direct, quantitative measure of how well an organization mitigates phishing campaigns through AI, SOAR, and SOC orchestration.

The End of Negative Metrics: From Human Blame to Systemic Resilience

For years, cybersecurity awareness programs have relied on the Click-Through Rate (CTR) as the central metric of phishing resilience.

CTR measures how many users click on a simulated phishing link — a negative-based indicator, built on the idea that fewer clicks mean higher awareness. While CTR once offered a simple behavioral snapshot, it has an inherent flaw: it measures failure, not success. It turns cybersecurity into a blame game, where every click is a mark of shame rather than a learning opportunity.

This model puts emotional pressure on employees, stigmatizes honest mistakes, and cultivates a defensive culture rather than a collaborative one.

In the new era of AI-assisted defense, this mindset is outdated. Modern organizations rely on AI-powered detection, automated containment, and human-AI collaboration.

The focus must shift from “Who clicked?” to “How fast did we detect and neutralize the threat?”

The true measure of maturity is no longer how well humans avoid being tricked, but how effectively machines and humans together detect, isolate, and neutralize phishing attacks. Thus, the Automated Resilience Index (ARI) represents a positive, system-based approach to measurement.

Each detection or containment event — whether triggered by AI or a vigilant user — is treated as a success signal, not a failure.

Instead of punishing clicks, ARI rewards detection speed and automation efficiency.

It celebrates the joint performance of humans and AI, reflecting a culture of empowerment rather than blame.

Background and Motivation

The CTR Metrics is completely obsolete.

For years, organizations have relied on phishing simulation click-through rates (CTR) to measure awareness. However, several issues undermine the reliability of CTR as a risk indicator:

- **False data from automated systems:** Email firewalls and sandboxing tools may open links automatically, creating false clicks.
- **Automated clicks:** JavaScript or embedded scripts may perform a click without any user interaction.
- **Awareness fatigue:** Users mechanically click “report” or “ignore” without real engagement.
- **Lack of context:** CTR measures user behavior, not system performance.
- **Delayed insight:** It provides retrospective results rather than predictive or operational metrics.

The result: CTR values no longer correlate with real organizational resilience.

The Shift to System-Level Metrics

As enterprises transition to AI-driven security orchestration (EDR/XDR/SOAR), the critical question changes from “Do users click?” to “How fast and autonomously is the threat neutralized?”

The key performance indicator for phishing defense becomes <u>time and automation</u>, not human reaction.

Introducing the Automated Resilience Index (ARI)

The Automated Resilience Index (ARI) quantifies the true capacity of a cybersecurity ecosystem to neutralize phishing or malicious email threats.
It integrates both the speed of neutralization and the degree of automation and success achieved during that process.

Dimension	Description	Desired Direction
Speed (MTTN)	**M**ean **T**ime **T**o **N**eutralization: how fast the threat is neutralized from detection to final remediation.	Lower is better (min value=0)
Containment Success (CSR)	Percentage of phishing or malicious items fully neutralized before user exposure.	Higher is better
Automation Efficiency (ACE)	Degree of automation: fraction of mitigation steps performed autonomously without human intervention.	Higher is better

Mathematical Foundation:

Formula derived from the 3 dimensions:

$$\boldsymbol{ARI = \pi \frac{CSR \ \cdot ACE}{log(MTTN + 1)}}$$

where:

π = 3.1416 (scaling constant)
CSR = Containment Success Rate (0–1)
ACE = Autonomous Containment Efficiency (0–1)
MTTN = Mean Time To Neutralization in seconds

Rationale of the Formula:
The formula is intentionally simple and physically meaningful:

- π ensures the upper bound (~10) aligns with human perception scales (similar to CVSS).
- CSR × ACE reflects the combined performance of automation and containment.
- 1 / log(MTTN) models diminishing returns: accelerating from 60 s → 10 s matters more than from 10 → 1 s.
- The logarithmic denominator reflects the temporal resistance of the environment — how long a phishing campaign remains active before neutralization.

The Meaning of Each Variable

CSR — Containment Success Rate

Percentage of detected phishing items successfully blocked, deleted, or quarantined before end-user access (value from 0 to 1).

Calculation Example:

$$CSR = \frac{Nbr\ of\ Email\ Neutralized\ Before\ Delivery}{Total\ Phishing\ Email\ Detected}$$

ACE — Autonomous Containment Efficiency

Proportion of mitigation actions completed automatically (via AI, playbooks, or rules) without human approval (value from 0 to 1).

Example:

If 80% of malicious emails are neutralized by a Tool (e.g. Defender for Office 365) automatically, and 20% required analyst validation or Human feedback:

$$ACE = 0.8$$

MTTN — Mean Time To Neutralization

Average duration (in seconds) from initial detection to full neutralization of a phishing or malicious email.

This can be measured using SIEM/SOAR logs:

- Detection timestamp (Defender, Proofpoint, etc.)
- Final mitigation timestamp (isolation/quarantine/deletion)

$$MTTN = \frac{\sum(t_{neutralized}\ \ t_{detected})}{n}$$

Interpreting ARI

ARI	Interpretation	Description
9-10	Exceptional	AI neutralizes threats nearly instantly (<1 s)
7-8	Excellent	Response within Seconds
5-6	Good	Neutralization within 1-2 minutes
3-4	Moderate	Delay under 5 minutes
1-2	Weak	Several minutes to hour
<1	Critical	Users likely clicked before Response

Rationale

User Click Behavior Data: statistics show that 10–15% of phishing clicks occur within the first minute of delivery, and 50% occur within five minutes.

Conclusions

The Automated Resilience Index (ARI) redefines how organizations measure phishing defense maturity.
Instead of judging users by CTR, ARI evaluates the cyberimmune system — the AI, automation, and orchestration that determine whether a phishing threat ever reaches an end user.

Key takeaways:

- CTR is obsolete in automated environments.
- ARI measures resilience, not behavior.
- Speed and automation define protection, not awareness.
- The logarithmic inverse-time model provides a stable, universally comparable scale.
- π-based calibration gives ARI mathematical elegance and interpretability.

Author's Note

This white paper introduces a quantitative model born from practical SOC operations and CISO-level strategy based on the large number of incidents analyzed in the last years — a bridge between behavioral analytics and technical automation performance.

The Automated Resilience Index (ARI) is designed for the next generation of cybersecurity leadership: measurable, scalable, and automation aware.
Author: Gianclaudio Moresi, 6340 Baar (Switzerland)

Key References:

Microsoft Digital Defense Report 2024

https://www.microsoft.com/en-us/security/security-insider/threat-landscape/microsoft-digital-defense-report-2024

The Quiet Evolution of Phishing – Microsoft Security Blog

https://www.microsoft.com/en-us/security/blog/2019/12/11/the-quiet-evolution-of-phishing

Verizon Data Breach Investigations Report (DBIR) 2024

https://www.verizon.com/business/resources/reports/dbir/

Google TAG – Trends in Phishing and Social Engineering 2024

https://blog.google/threat-analysis-group/trends-in-phishing-and-social-engineering-2024/

Research & Whitepapers

Threat-Informed Cyber Resilience Index: A Probabilistic Quantitative Approach to Measure Defence Effectiveness Against Cyber Attacks

https://arxiv.org/abs/2406.19374

Cyber Resiliency Metrics, Measures of Effectiveness, and Scoring – MITRE

https://www.mitre.org/sites/default/files/2021-11/prs-18-2579-cyber-resiliency-metrics-measures-of-effectiveness-and-scoring.pdf

Quantitative Measurement of Cyber Resilience: Modeling and Experimentation

https://arxiv.org/abs/2303.16307

Resilience Metrics for Cyber Systems – Igor Linkov, Julia Allen et al.

https://www.researchgate.net/publication/263176904_Resilience_metrics_for_cyber_systems