FRAMEWORK · DERIVED FROM THREAT TERMINAL

The Action Pause

A ten-second habit for anything that asks you to act. A post-filter, user-layer defense for the phishing that now slips through every technical control.

By Scott Altiparmak · April 2026 · Calibrated against 2,511 classifications from the Threat Terminal study

Empirical findings (DOI) →

Read the essay

60.5%

of phishing misses occurred at the highest confidence level

Overconfidence, not guessing, is the dominant failure mode.

86.8% → 76.0%

detection drop when phishing displayed passing authentication

Clean SPF/DKIM did not protect users. It quieted them.

20.5%

miss rate on authority impersonation

The highest-miss technique, and barely addressed by standard training.

12.5%

miss rate on credential harvest

The category most curricula drill. Users already detect it reliably.

THE PROBLEM THIS ADDRESSES

Modern email security handles most inbound phishing at the technical layer. Secure email gateways, native filtering from Microsoft and Google, and authentication standards (SPF, DKIM, DMARC) do what they can do. What remains is, by selection, the hardest class to detect: clean authentication, clean infrastructure, fluent prose, and well-targeted context. This is the residual class, and it is where real breaches originate.

Existing user training was built against pre-AI phishing with detectable indicators: misspellings, awkward tone, suspicious sender names. Those indicators do not exist in the residual class. The Threat Terminal study demonstrated this directly: detection of phishing with passing authentication dropped from 86.8% to 76.0%. Clean signal did not make users safer. It quieted them.

The Action Pause is designed for that class, and only that class. It does not replace filters. It addresses the specific failure mode that filters cannot close and indicator-based training does not touch.

TEN SECONDS. EVERY TIME.

The Action Pause

TRIGGER

“This is asking me to do something.”

Click · Sign in · Approve · Pay · Forward · Reply with information · Install · Scan a QR code

ASK THREE QUESTIONS

01 Did I expect this?

02 Is this the normal channel for this request?

03 If it is fake, what breaks?

CALIBRATE HONESTLY

If you are not certain, or certain but unverified on anything consequential, verify through a separate channel before you act.

STANDING RULE

Clean authentication is not safety. It is silence.

For any consequential action, the email itself is never sufficient evidence. Verify through a known phone number, a known portal, or in-person, somewhere the attacker cannot impersonate.

WHY THE TRIGGER MUST BE REFLEXIVE

Phishing does not usually succeed because the user missed a signal. It succeeds because the user noticed something was off but had already acted by the time the noticing caught up. “I knew it was weird the second I hit the button” is the modal account of a real-world phishing loss. The signal arrives. The reflection never fires in time to stop the action.

This is a cognitive property of a well-crafted phish, not a moral failing of the target. The message keeps the recipient in the action loop (reading, recognizing, responding) without surfacing the reflection that would interrupt it.

The Action Pause only works if its trigger fires **before** the action. “Pause when you notice an ask” helps only if noticing an ask is itself reflexive, surfacing automatically rather than requiring the user to remember a framework in the moment. The three questions and the calibration rule are deliberate interventions. They only work if a reflexive trigger hands them the moment.

That is what good training builds: a reflex that fires on **“this is asking me to act.”** The Threat Terminal data shows this shape directly. Within a single session, accuracy rose from 79.3% on the first card to 89.3% by the fifth. That is not knowledge acquisition in five minutes. It is a trigger getting faster. Annual awareness modules do not produce this pattern. Short, frequent exposures do, which is why the curriculum principle below is load-bearing rather than stylistic.

WHY EACH QUESTION IS IN THE FRAMEWORK

Each question maps to a measured failure mode. The question set is not a generic mnemonic; it is weighted to the categories where the data shows users actually miss.

> Q01 **Did I expect this?**
>
> Addresses: Hyper-personalization (14.5%), fluent prose (16.2%)
>
> AI-generated content reads fluently and targets your role. The defense is not better reading. It is knowing whether this request was expected in the first place.

Q02 Is this the normal channel for this request?

Addresses: Authority impersonation (20.5%), pretexting (16.3%)

Impersonation works by mimicking a plausible sender. Knowing what plausible actually looks like, how this person or system usually reaches you, is the defense.

Q03 If it is fake, what breaks?

Addresses: Urgency (16.8%), and every action with real blast radius

Urgency moves decisions out of the consequence frame. This question pulls them back. Money, credentials, access, or data change the verification bar.

CAL Calibration rule

Addresses: 60.5% of misses at highest confidence

The most common failure pattern in the data is not uncertainty that went wrong. It is certainty that went wrong. The calibration rule does not ask users to be more skeptical in the abstract. It asks them to treat unverified certainty, on consequential actions, as a warning sign about themselves.

HOW THIS DIFFERS FROM PRIOR FRAMEWORKS

The Action Pause builds on a body of prior work rather than replacing it. What is new is the trigger, which is structural rather than content-based, and the calibration rule. Both are responses to measured failure modes that did not exist when earlier frameworks were written.

FRAMEWORK	TRIGGER	CALIBRATION	SCOPE
Stop. Think. Connect. NCSA / DHS, 2010	Any online action or suspicious content	Implicit	General online safety
NIST SP 800-50 / SP 800-16 NIST awareness guidance	Curriculum specification	Not addressed	Enterprise awareness programs

PhishGuru embedded training Kumaraguru et al., CMU	Post-click in simulation	Not addressed	Simulation-tethered intervention
Indicator-based heuristics Vendor and SANS curricula	Email content cues (URLs, headers, grammar)	Not addressed	Pre-AI phishing
The Action Pause This work	Request structure (any action ask)	Explicit rule: unverified certainty = warning	Post-filter, AI-era residual phishing

Earlier frameworks remain useful for the attacks they were designed against. The Action Pause is specifically for the residual class that passes modern filters. Content-level heuristics no longer isolate that class, which is why a structural trigger is required.

FOR PRACTITIONERS: CURRICULUM PRINCIPLES

Four principles for building awareness programs around the Action Pause. Each is tied to a specific pattern in the Threat Terminal data, not to pedagogical preference.

01 Teach the trigger first.

Most misses happen because the user did not notice they were about to act. Recognizing an action request is the prerequisite skill. The questions come second and are easier to teach once the trigger is reflexive.

02 Reweight toward the high-miss categories.

Shift curriculum time from credential harvest drills (12.5% miss, already solved by training) toward authority impersonation (20.5%), urgency (16.8%), pretexting (16.3%), and fluent prose (16.2%). These are where detection actually fails.

03 Short and frequent, not annual and long.

Session-level accuracy in the study rose from 80.2% (Session 1) to 88.6% (Session 3). Within-session, first-card accuracy was 79.3% versus 89.3% by the fifth. Iterative exposure does real work. Sixty-minute annual modules do not replicate this pattern.

04 Surface overconfident misses by name.

The highest-leverage feedback event is not "you got one wrong." It is "you were certain, and you were wrong." Name the miscalibration directly. The 60.5% overconfidence rate is the behavior this feedback pattern targets.

PILOT AVAILABILITY

The Action Pause is designed for enterprise validation. I am preparing to run it through a gamified training deployment built on the Threat Terminal research instrument, with the same measurement discipline applied to training outcomes: per-technique detection rates, confidence calibration, and out-of-band verification behavior.

Organizations interested in piloting the framework, or researchers interested in the dataset and methodology, are welcome to reach out.

HOW TO CITE

```
Altiparmak, S. (2026). The Action Pause: a post-filter, user-layer defense against AI-generated phishing. Derived from the Threat Terminal study (doi:10.5281/zenodo.19410549). scottaltiparmak.com/research/action-pause
```

EMPIRICAL BASIS doi.org/10.5281/zenodo.19410549 (preliminary empirical findings)

STUDY PROTOCOL doi.org/10.5281/zenodo.19059296

The Action Pause builds on prior work including SANS / NCSA Stop. Think. Connect., NIST SP 800-50 / 800-16 awareness guidance, Kumaraguru et al.'s PhishGuru embedded training (Carnegie Mellon), Wash's research on expert detection heuristics, and Canfield, Fischhoff, and Davis's work on phishing confidence calibration. The novel contribution of this framework is the structural trigger and the explicit calibration rule, both calibrated against measured miss rates in a controlled dataset of AI-generated stimuli.