From Framework to Operations:
The CPF Deployment-Validation Loop

A Preprint

Giuseppe Canale, CISSP
Independent Researcher
g.canale@cpf3.org
cpf3.org
ORCID: 0009-0007-3263-6897

February 2026

Abstract

The Cybersecurity Psychology Framework (CPF) defines 100 pre-cognitive vulnerability indicators across 10 categories. The Validation Roadmap [4] provides a four-tier methodology for empirically testing these indicators. But a structural problem remains: Tier 4—the only tier that provides operational validity—requires live SOC deployment, and organizations will not deploy a framework that has not yet demonstrated operational value. This is a chicken-and-egg problem inherent to any predictive framework that targets phenomena not observable through controlled experimentation alone.

This paper resolves the paradox. We show that when CPF is deployed as an additive layer—alongside, not replacing, existing detection tools—the act of deployment itself becomes the validation mechanism. Every day of operation simultaneously produces operational intelligence and validation evidence from the same data stream. We formalize this as the deployment-validation loop, define the metrics that serve both purposes, and present a five-level maturity model that describes how an organization's CPF capability evolves as evidence accumulates.

Keywords: deployment-validation loop, additive detection, SOC integration, maturity model, pre-cognitive vulnerability, operational validation

1 Introduction

Security Operations Centers rely on a layered detection architecture: SIEM platforms aggregate log data, EDR tools monitor endpoint behavior, and UEBA systems profile user activity to identify anomalies. Each layer targets a different signal. None of them targets the psychological states that precede security-relevant human decisions.

The CPF [1] proposes to fill this gap. It maps 100 pre-cognitive vulnerability indicators—authority susceptibility, cognitive overload, social influence, affective states, and others—to a detection taxonomy that can be monitored continuously. The Implementation Companion [3] provides the technical machinery: the OFTLISRV detection schema, the Bayesian indicator network, and the Convergence Index that aggregates individual indicator states into an organizational risk signal.

The Validation Roadmap [4] provides the empirical methodology: four tiers of testing, from LLM proxy validation through live operational deployment, designed to build convergent evidence across independent methods. The roadmap is rigorous and executable. But it contains a structural assumption that this paper challenges: that validation and deployment are sequential stages, with deployment coming after sufficient validation has accumulated.

For CPF, this assumption creates a deadlock. The framework targets pre-cognitive processes [7]—states that, by definition, cannot be fully captured in controlled experimental settings [5]. The only validation method that provides genuine operational evidence is Tier 4: prospective prediction within a live SOC. But Tier 4 requires deployment. And deployment, under the sequential model, requires prior validation. The framework cannot be adopted because it cannot be validated, and it cannot be validated because it cannot be adopted.

This paper shows that the deadlock is an artifact of the sequential assumption, not an inherent property of the problem. When CPF enters a SOC as an additive layer—one signal among many, never replacing existing tools—the risk calculus changes fundamentally. Deployment becomes costless in operational terms, and the data it generates serves validation and operations simultaneously.

1.1 Scope

This paper assumes familiarity with the CPF suite: the framework [1], the SiliconPsyche protocol [2], the Implementation Companion [3], and the Validation Roadmap [4]. It does not restate the indicator taxonomy, the detection logic, or the validation tier methodology. Its contribution is the argument that deployment need not wait for validation, and the formal description of how the two processes merge into a single loop.

2 The Adoption Paradox

2.1 The Standard Model

The standard model for adopting a new detection capability follows a linear sequence: the capability is proposed, tested in controlled conditions, validated against known outcomes, and then deployed into production. This model works well for detection tools that target observable phenomena—network intrusions, known malware signatures, anomalous login patterns. The phenomena can be reproduced in test environments, the tool's performance can be measured against ground truth, and the decision to deploy is made on the basis of demonstrated efficacy.

2.2 Why It Fails for Pre-Cognitive Frameworks

CPF targets processes that operate below conscious awareness [6]. This creates three properties that break the standard model.

The phenomena cannot be reliably reproduced in isolation. Pre-cognitive vulnerability states emerge from the interaction of psychological predisposition, organizational context, and temporal conditions. A controlled experiment can approximate individual factors, but the convergent states that drive actual security incidents require the full complexity of a live operational environment. This is precisely why the Validation Roadmap identifies Tier 4 as the tier that provides operational validity—and precisely why Tier 4 requires live deployment.

The measurement instrument changes what it measures. Any experimental setup that makes subjects aware of psychological vulnerability testing activates the Hawthorne effect [5], suppressing the pre-cognitive processes that CPF targets. Tier 3 of the Validation Roadmap mitigates this through differential awareness design, but mitigation is not elimination. The only context in which pre-cognitive vulnerability states operate without experimental artifact is normal, unstructured organizational activity.

Validation evidence accumulates, it does not arrive. There is no single experiment that can prove CPF works. The Validation Roadmap explicitly acknowledges this: full validation requires convergence across four independent methods. This means that validation is not a gate that opens before deployment—it is a process that runs continuously and improves over time.

2.3 The Paradox Stated Formally

Let V represent the set of validation evidence required for adoption, and D represent the deployment that generates operational validation data. Under the standard model, D is contingent on V: deployment requires prior validation. But for CPF, a necessary subset of V—specifically, Tier 4 operational evidence—is contingent on D: that evidence can only be generated through deployment. Therefore:

$$D \text{ requires } V, \quad V \text{ requires } D$$

This is the adoption paradox. It is not unique to CPF—it applies to any predictive framework whose primary validation method is prospective operational testing. But it is particularly acute for pre-cognitive frameworks because the other validation tiers (controlled experiments, retrospective analysis) cannot fully substitute for operational evidence.

3 Why Additive Deployment Resolves the Paradox

The paradox assumes that deployment carries operational risk: if the framework is wrong, deploying it will cause harm. This assumption is valid when a new tool replaces an existing one. It is not valid when the new tool adds a layer without displacing anything.

3.1 The Additive Layer Property

CPF is designed to operate as an additive detection layer. It does not replace SIEM, UEBA, or EDR. It does not alter the SOC's incident response procedures. It does not trigger automated actions. It produces a signal—the state of the 100 indicators and the Convergence Index—that exists alongside the signals already present in the SOC environment.

This means that if CPF's predictions are wrong, the consequence is a signal that does not correlate with incidents. The SOC continues to function exactly as it did before CPF was deployed. The existing tools remain authoritative. The only cost is the analyst's time spent evaluating a signal that proved uninformative—and even this cost is bounded, because the deployment-validation loop (described in Section 4) makes uninformative signals visible within weeks, not months.

3.2 The Dual-Purpose Data Stream

When CPF operates in a live SOC, every operational period produces two types of information simultaneously. First, *operational intelligence*: which indicators are elevated, whether the Convergence Index suggests heightened organizational risk, and where an analyst should direct attention. Second, *validation evidence*: whether the predictions made at the start of the period were followed by actual incidents, how the lead time compared to existing tools, and whether the Convergence Index correlated with incident severity.

These two types of information come from the same data stream. There is no additional instrumentation, no separate data collection, no experimental overhead. The act of running CPF in the SOC produces both outputs as a natural consequence of its operation. This is the property that dissolves the paradox: deployment does not need to wait for validation because deployment is how validation happens.

3.3 Risk Asymmetry

The decision to deploy CPF as an additive layer presents a risk asymmetry that favors early deployment. If CPF adds value, the SOC gains a capability that existing tools do not provide: early warning of psychologically-driven vulnerability states. If CPF does not add value, the SOC loses nothing—the layer produces uninformative signals that are recognized as such within the first operational period and can be deprioritized or removed. The expected cost of early deployment is bounded and small. The expected cost of delayed deployment—in the event that CPF does add value—is the missed early-warning capability for the duration of the delay.

4 The Deployment-Validation Loop

4.1 Loop Structure

The deployment-validation loop operates on a repeating cycle. Each cycle consists of four phases:

Phase 1: Prediction. At the start of each operational period (default: 24 hours), the CPF engine evaluates all 100 indicators based on current organizational data and produces a state vector: the Green/Yellow/Red classification of each indicator, the Convergence Index value, and the set of categories in elevated state. This prediction is sealed with a cryptographic timestamp before any incident data for the period is available, following the prediction-first protocol defined in the Validation Roadmap [4].

Phase 2: Operation. The SOC operates normally. The CPF state vector is available to analysts alongside other detection signals. No special procedures are required. Analysts treat the CPF signal as they would any other indicator in their environment: investigate if it correlates with other signals, note it if it does not.

Phase 3: Outcome. At the end of the operational period, actual security events are recorded. This data comes from the SOC's existing incident management system—no additional data collection is needed.

Phase 4: Evaluation. The sealed prediction from Phase 1 is compared against the outcomes from Phase 3. Four metrics are computed: precision (fraction of elevated-risk predictions followed by incidents), recall (fraction of incidents preceded by elevated-risk predictions), lead time (how far in advance the elevated state appeared), and Convergence Index correlation with incident severity. These metrics are logged and accumulated across periods.

The loop then repeats. Over time, the accumulated metrics produce a clear picture: CPF either adds predictive value beyond existing tools, or it does not.

4.2 Metrics as Dual-Purpose Instruments

The four metrics defined in the loop serve both operational and validation purposes simultaneously:

An analyst uses precision to calibrate how much weight to give the CPF signal relative to other indicators. A researcher uses the same precision value to assess whether the framework's predictions are statistically meaningful. The metric is identical in both cases. No instrumentation is duplicated.

4.3 Self-Correcting Dynamics

The loop is self-correcting by design. If CPF's predictions consistently fail to precede incidents (low recall) or consistently predict incidents that do not occur (low precision), this pattern becomes visible in the accumulated metrics within the first 30 days of operation. An analyst observing this pattern will naturally reduce the weight assigned to the CPF signal. No formal decision is required—the loop adjusts organically through the SOC's normal operational behavior.

Conversely, if CPF consistently identifies elevated risk states that precede incidents not flagged by other tools, analysts will notice and increase their attention to the signal. This is exactly the behavior that a useful additive layer should produce.

5 Relationship to the Validation Roadmap

This paper does not supersede the Validation Roadmap [4]. The Roadmap defines *what* must be tested and *how* at each tier. This paper adds a third dimension: *when.* The answer is: immediately, in production, for Tier 4—and in parallel with production deployment, for Tiers 1 and 2.

5.1 Tier Parallelization

The Validation Roadmap presents four tiers in a sequence that reflects epistemic dependency: each tier answers a progressively stronger question about the framework's validity. This sequence is appropriate for building the scientific case. It is not required for deployment.

Tier 1 (LLM proxy testing) and Tier 2 (retrospective correlation) can run in parallel with live deployment. They do not depend on deployment data, and they do not block it. Their results strengthen the scientific case over time but are not prerequisites for the SOC to begin generating operational experience with CPF.

Tier 3 (controlled human experiments) remains the most resource-intensive tier and is appropriately deferred until organizational partnership and IRB approval are secured. Its results inform the framework's refinement but do not gate deployment.

Tier 4 in the Roadmap is, in effect, this paper's deployment-validation loop. The Roadmap defines it as the final tier. This paper argues that it can be the first.

5.2 Convergence Across Methods

Running the deployment-validation loop in parallel with Tiers 1 and 2 produces an important benefit: early convergence signals. If Tier 1 shows that a specific indicator is robust across LLM architectures, and the deployment loop simultaneously shows that the same indicator correlates with actual incidents, the convergence between synthetic and operational evidence arrives months earlier than it would under sequential execution. The scientific case strengthens continuously rather than building in discrete stages.

6 The CPF Organizational Maturity Model

As the deployment-validation loop runs, an organization's relationship with CPF evolves. The maturity model describes this evolution as five emergent levels, each defined by the pattern of evidence the loop has produced—not by completion of external validation steps.

6.1 Level 1: Unaware

The organization has no CPF deployment. Security incidents driven by psychological factors are discovered after the fact, if they are attributed to human behavior at all. This is the baseline state for the majority of SOCs today.

6.2 Level 2: Monitoring

CPF has been deployed as an additive layer. The deployment-validation loop is active. Metrics are accumulating but have not yet produced a clear signal—precision, recall, and lead time are still within the range of noise. Analysts are aware that the CPF signal exists but have not yet formed a reliable judgment about its value. This is the expected state for the first 30–60 days of operation.

6.3 Level 3: Predictive

The accumulated metrics show a consistent pattern: the CPF Convergence Index and elevated indicator states appear before incidents with a lead time of at least 24 hours in a statistically meaningful fraction of cases. The signal has proven itself to be informative. Analysts treat it as a credible early-warning indicator, though they have not yet begun to act on it proactively.

The transition from Level 2 to Level 3 is the critical inflection point. It is the moment when CPF moves from being an experiment to being a tool.

6.4 Level 4: Proactive

Analysts have begun to act on elevated CPF states before incidents occur. When the Convergence Index rises, targeted investigation or preventive measures are initiated without waiting for a triggering event from other detection tools. The organization is using CPF not just to understand what happened, but to intervene before it does.

This level requires organizational process change: the SOC's response procedures must include CPF signals as legitimate triggers for investigation. It also requires analyst confidence built through the Level 2–3 experience—analysts will not act on a signal they do not trust.

6.5 Level 5: Optimized

The deployment-validation loop has produced sufficient data to identify which indicators and indicator combinations are most predictive in this specific organizational context. Indicator weights are adjusted to reflect observed correlations rather than theoretical priors. The Convergence Index is calibrated to the organization's baseline incident rate. The feedback loop between operational outcomes and framework parameters is active and continuous.

At this level, CPF has become a self-improving detection capability. Each incident—whether predicted or not—provides data that refines the model's accuracy for future predictions.

7 What Success and Failure Look Like

A critical property of the deployment-validation loop is that both success and failure produce useful information. This distinguishes it from validation approaches that treat negative results as dead ends.

7.1 Success Signature

Success manifests as a progressive divergence between CPF metrics and the null baseline. Specifically:

Precision rises above the null model (random prediction at the base rate of incidents) within the first 60 days and remains above it. Recall exceeds the detection rate of UEBA systems on human-behavior-driven incidents. Lead time becomes consistently positive—the elevated state appears before the incident, not after. Convergence Index correlation with incident severity exceeds $r = 0.3$ and trends upward as the indicator model is refined.

When these signatures appear together, the deployment-validation loop has produced the same evidence that Tier 4 of the Validation Roadmap defines as operational validity. The loop has validated the framework through use.

7.2 Failure Signature

Failure manifests as convergence with the null baseline. Precision does not exceed the base rate. Recall does not exceed what UEBA already provides. Lead time is zero or negative—the CPF signal appears at the same time as or after the incident. The Convergence Index does not correlate with incident severity.

This is not a catastrophic outcome. It is an informative one. It tells the organization that the framework, as currently configured, does not add predictive value in their specific environment. It tells the CPF development effort that specific indicators or categories need revision. And it costs the organization nothing operationally, because the additive layer property meant that existing detection capabilities were never compromised.

7.3 Partial Results

The most likely near-term outcome is neither full success nor full failure, but a partial signal: some categories of indicators prove predictive while others do not. Authority-based vulnerabilities ([1.x]) and cognitive overload ([5.x]) may produce strong signals in a SOC environment, while social influence ([3.x]) or group dynamics ([6.x]) may not—perhaps because the organizational context does not activate those vulnerability patterns at observable frequency.

Partial results are the expected outcome of any framework operating in a specific organizational context. They are not a failure of the framework. They are information about where the framework adds the most value in that context, and they drive the Level 5 optimization process described in Section 6.5.

8 Conclusion

The adoption paradox for pre-cognitive detection frameworks is real: you cannot validate what you do not deploy, and you cannot deploy what you have not validated. But the paradox dissolves when deployment is additive. A framework that adds a signal without displacing any existing capability carries no operational risk. Its first day of operation produces both intelligence and evidence. The SOC itself becomes the laboratory.

The deployment-validation loop described in this paper is not a shortcut around rigorous validation. It is the most rigorous form of validation available for a framework that targets phenomena only observable in authentic operational conditions. The Validation Roadmap's other tiers—LLM proxy testing, retrospective correlation, controlled experiments—provide complementary evidence that strengthens the scientific case. But they cannot substitute for the evidence that only live deployment produces.

The maturity model shows that an organization's CPF capability does not need to be fully validated before it becomes useful. It evolves through use: from monitoring to prediction to proactive intervention to self-optimization. At each stage, the evidence that justifies the next stage is generated by the current stage itself.

The loop closes. The gap between theory and practice is not bridged by a paper or a proof—it is bridged by operation.

References

[1] Canale, G. (2025). The Cybersecurity Psychology Framework: A Pre-Cognitive Vulnerability Assessment Model. *CPF Technical Report Series.*

[2] Canale, G. and Thimmaraju, K. (2026). The Silicon Psyche: Anthropomorphic Vulnerabilities in Large Language Models. *CPF Technical Report Series*, Version 1 (Revision 11).

[3] Canale, G. (2025). Operationalizing the Cybersecurity Psychology Framework: A Systematic Implementation Methodology. *CPF Technical Report Series.*

[4] Canale, G. (2026). Toward Empirical Validation of the Cybersecurity Psychology Framework: A Tiered Methodological Roadmap. *CPF Technical Report Series.*

[5] Landsberger, H. A. (1958). *Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.* Ithaca: Cornell University Press.

[6] Kahneman, D. (2011). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

[7] Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity. *Brain*, 106(3), 623-642.