The Promotion Gap. Why a public mitigation remains an artifact until tested on the actual target

Article VI. The CVE-2026-31431 (Copy Fail) case on WSL2 kernel 6.6.87.2 as empirical terrain for the promotion gate between technical artifact and normative dependency.

A public security recommendation is an artifact whose tested scope does not mechanically coincide with the scope of the industrial targets to which it is prescribed. Crossing the promotion gate toward normative dependency cannot be delegated to the issuing authority. Demonstration in four moves, from a case observed on April 30, 2026.

Key points:

> *"A CERT-EU mitigation is an artifact. Its promotion to normative dependency goes through behavioral validation on the exact target."*
>
> *"The promotion gap denotes the measurable distance between the scope tested by the authority and the scope prescribed to industry."*
>
> *"On WSL2 kernel 6.6.87.2, four public mitigations were behaviorally invalidated within the same operational session."*
>
> *"The effective trace in the kernel sense, accessible via `/proc/<pid>/status`, is the only invariant that distinguishes effective mitigation from cosmetic mitigation."*

Introduction

Article V of this series posed the hexagonal architecture as a governance frame for training phases of AI applied to regulated domains. It introduced three operational concepts: the *promotion gate* as the boundary between experimental artifact and normative dependency, *traceable deliberation* as the documented procedure for crossing it, and *genealogical debt* as the liability that any artifact carries with it when promoted without having traced the conditions of the passage.

These concepts were formulated from the ToxTwin experimental terrain, on an internal trajectory, under the author's control. They deserved a test on an external terrain, where the issuing authority is not the author, where the cadence of application is imposed by the community, and where the conditions of validation are not negotiable.

That terrain presented itself on April 30, 2026, in the form of CVE-2026-31431 known as Copy Fail, a local privilege escalation in the Linux crypto subsystem exploitable via 732 bytes of Python from any unprivileged account.

The thesis fits in one sentence. **A public security mitigation remains an artifact until it has been behaviorally tested on the exact target where it is prescribed, and crossing the promotion gate toward normative dependency cannot be delegated to the issuing authority.**

The domain of validity is circumscribed. The argument concerns mitigation recommendations issued by public cybersecurity authorities applied to heterogeneous production environments, and it is demonstrated on the particular case of the WSL2 kernel 6.6.87.2-microsoft-standard-WSL2 facing CERT-EU and openwall recommendations for Copy Fail. It generalizes to public mitigations applied to any kernel variant not explicitly tested by the issuing authority. It concerns neither the general conduct of information security, nor patch management doctrine, which have their own established frames. The demonstration follows four moves: the CERT-EU artifact within its claimed scope, the empirical proof of its inoperancy outside that scope, the alternative mitigation validated on the actual target, and the promotion criterion that follows.

Move 1. The CERT-EU artifact and its tested scope

On April 29, 2026, the Theori team published the coordinated disclosure of CVE-2026-31431. An optimization introduced in 2017 in the `algif_aead.c` module of the Linux crypto subsystem allowed an unprivileged user to write four controlled bytes into the page cache of a readable file, by combining the AF_ALG socket, the splice() primitive, and the peculiarities of the AEAD authencesn template. Modifying the page cache of a setuid binary suffices to gain root on the next invocation. The public proof-of-concept, distributed as a 732-byte Python script requiring neither kernel offset nor race window, ran on Ubuntu 24.04 LTS, Amazon Linux 2023, RHEL 10.1, and SUSE 16. The CVSS score of 7.8 likely understated the operational impact, whose grey market typically recognizes the value of a Parisian apartment.

During April 29, several authorities published their mitigation recommendations while waiting for distribution kernel patches. CERT-EU, openwall, and the kernel.org community converged on the same instruction: neutralize the `algif_aead` module via a `modprobe install algif_aead /bin/false` directive, accompanied by an immediate `rmmod algif_aead` to clear memory. The recommendation was technically sound. It aimed to render unavailable the crypto handler that exposed the write primitive, without touching userspace services that use the upper layers (dm-crypt, kTLS, IPsec, OpenSSL).

The tested scope of this recommendation, as it could be reconstituted from the published advisories, comprised the distributions explicitly listed in the Theori disclosure and a plausible set of standard configurations on official distribution kernels. The authority did not claim universal validation. It could not, and that is rational. No public cybersecurity authority has the means to test a recommendation on

the combinatorial set of kernel, distribution, version, and architecture variants deployed in industry. Authorities test on dominant targets, publish on dominant targets, and leave each organization the responsibility of verifying portability on peripheral variants.

The prescribed scope, however, was universal. The CERT-EU recommendation was formulated without qualification, applicable to any Linux distribution shipped since 2017. This asymmetry between bounded tested scope and unbounded prescribed scope is the very condition of the public artifact. It is not a defect of the issuing authority, it is its operating condition at scale. But it has a direct operational consequence: the recommendation prescribed to a heterogeneous industrial estate necessarily contains a non-zero promotion gap whose amplitude is known to no one at the moment of application. The implicit doctrine holds that the authority publishes, that industry applies, and that silence on variants means they hold. On that last point, the implicit doctrine is wrong more often than is generally admitted.

Move 2. The empirical proof of inoperancy on WSL2 kernel 6.6.87.2

The environment of a Windows Server 2025 with WSL2 on which the investigation was conducted runs Ubuntu 24.04 on WSL2 kernel 6.6.87.2-microsoft-standard-WSL2, hosting pseudonymized clinical data and an active ML stack. Seven local accounts coexist on it, six of them non-administrative. The profile corresponds exactly to the target the CERT-EU recommendation is supposed to cover: Ubuntu 24.04 LTS, multi-user, regulated data, urgency of application before kernel patch availability.

The recommendation was applied according to published instructions: creation of `/etc/modprobe.d/disable-algif-aead.conf` with the install directive, unloading of the module if loaded. A behavioral validation was conducted in parallel, consisting in reproducing the priming primitive of the exploit in a controlled Python script: opening an AF_ALG socket, binding on the `authencesn(hmac(sha256),cbc(aes))` template, measurement of the value returned by the kernel. The test invalidated the mitigation. The bind succeeded. The attack primitive remained available.

The technical cause was identified by instrumentation. The WSL2 kernel, on this precise build, uses a module loading path that calls `request_module()` on the kernel side when a userspace requests an AF_ALG crypto service, and that path bypasses the userspace modprobe resolution in which the install directive had been registered. The module is reloaded on demand, despite the blacklist.

Four successive variants of the recommendation were tested in the same operational session, each more restrictive than the previous one.

1. The blacklist of the complete af_alg framework, covering af_alg, algif_aead, algif_skcipher, algif_hash, and algif_rng, failed for the same reason.

2. The kernel-level deactivation of module loading via the sysctl `kernel.modules_disabled=1` produced a particular operational trap: the lock also forbids unloading, which froze in memory the modules already loaded by the previous tests.
3. Physical deletion of the `.ko` files was blocked by the sysctl lock in progress.
4. A fifth attempt, which would have required restarting the WSL2 virtual machine via a PowerShell command on the Windows host, failed for trivial human reasons whose detail I shall spare myself from documenting.

The brute empirical fact is that on WSL2 kernel 6.6.87.2, in the session of April 30, 2026, **four variants of the public recommendation were behaviorally invalidated by the same syscall test**. None neutralized the attack primitive. Without the behavioral validation conducted in parallel with the application, the incident would have been closed in four declarative compliance tickets each accumulating the satisfaction of having applied the official directive. The distance between *having applied* and *having mitigated* was, in this particular case, four iterations.

This distance is not anecdotal. It is the local measure of the promotion gap between the scope where the recommendation was tested by the authority and the scope where it is prescribed to industry. Microsoft announced in 2025 more than four million WSL2 instances in enterprise use. None of these environments figures in the official CERT-EU or openwall advisories. All are, at the moment of their application of the recommendation, in the situation where declarative compliance coincides with behavioral vulnerability. The probability that all present the same non-standard behavior as the one observed on 6.6.87.2 remains to be established. The probability that at least a non-negligible fraction present a comparable behavior is, however, very high. If the doctrine of public mitigation held silently across these four million variants, the convergence would be remarkable. It is not. It is merely assumed.

Move 3. The empirical alternative mitigation and its behavioral validation

Once it was acquired that the public recommendation did not hold on the actual target, two trajectories were open. The first consisted in waiting for the WSL2 kernel patch published by Microsoft, hence leaving the exposure window open for an uncontrolled duration, hardly compatible with the nature of the data hosted. The second consisted in designing an alternative mitigation from the available kernel layers, independently of the failing modprobe doctrine.

The retained path mobilizes the seccomp BPF mechanism, integrated into the Linux kernel since 3.5 and standardized by industry since Docker 1.10. A surgical BPF filter refuses the `socket(AF_ALG, ...)` syscall at the kernel level, returning EAFNOSUPPORT (errno 97), without touching any other vector. The priming primitive of Copy Fail is dead at entry. The filter is attached to the connection processes of unprivileged accounts via a shell-wrapper substituted in `/etc/passwd`, which makes the application irrevocable within a session. The administrator account is deliberately exempted for investigation and debug needs, creating a kernel-level asymmetry between user accounts and operational accounts.

The final behavioral validation rests on a reproducible command. For each unprivileged account, execution of a Python script that attempts the creation of the AF_ALG socket and captures the returned value. Six user accounts return PROTECTED (errno=97). The administrator account returns VULNERABLE, that is, a valid socket. The most direct proof of filter attachment goes through reading `/proc/<pid>/status`, where the invariant `Seccomp: 2` indicates active FILTER mode. This information is not a message printed by the shell. It is the state of the kernel describing itself. It cannot lie by quoting default or by silent redirection, which are the usual pathologies of performative traces produced by cosmetic validation scripts. The kernel does not lie. The shell, sometimes, does.

The mitigation holds. It was tested on the exact target, by a behavioral test that reproduces the attack primitive, and the effective trace in the kernel sense confirms filter attachment. No official advisory mentioned it on the date of the incident. It was produced by local empirical operation, in response to the documented inoperancy of public recommendations. It constitutes, by construction, an artifact whose tested scope is exactly the scope where it is applied. The promotion gate between artifact and normative dependency was crossed by the same operation that produced the artifact. This is what distinguishes this mitigation from public recommendations: it carries no residual promotion gap.

Move 4. The promotion criterion. The promotion gap as operational concept

The Copy Fail case makes explicit what remained implicit in Article V. The promotion gate is not only an internal governance procedure for artifacts produced inside the organization. It is a verificational dispositive whose necessity extends to any artifact received from outside, including when the outside is a public reference authority. This extension introduces a new concept, in direct continuation of the promotion gate, which I propose to call the *promotion gap*.

The promotion gap denotes the measurable distance between the scope where an artifact has been tested by its issuing authority and the scope where it is prescribed as normative dependency. This gap is not a defect of doctrine. It is its condition of existence. Any public recommendation has a non-zero promotion gap, because no authority can test its artifact on the combinatorial set of industrial targets. The operational question is not to reduce the gap to zero, which is impossible, but to render it measurable and attributable. *Measurable* means that the organization applying the recommendation can specify the distance between its own target and the targets tested by the authority, and that it conducts, for variants not covered, a behavioral validation whose result is documented. *Attributable* means that the responsibility for crossing the promotion gate is assigned to an actor with the technical and legal means to assume it. This responsibility cannot be delegated to the issuing authority, which has neither access to the target nor mandate to cover it. Nor can it be diluted in the audit chain, which measures only declarative compliance. It remains the responsibility of the organization that operates the environment, and of it alone.

The promotion criterion is then formulated as follows. An artifact received from outside is promoted to normative dependency for a given target only after production of a behavioral validation trace on that

exact target, attested by a kernel invariant or an observable equivalent of equivalent level. This formulation is demanding. It is so deliberately. It recognizes that declarative compliance remains useful for audit, without accepting that it substitute for security. It also recognizes that the systematic establishment of a behavioral validation file per variant represents an engineering expense that is not negligible, and that must be proportionate to the estimated promotion gap. For organizations whose information estate is homogeneous and standard, the gap is small and declarative compliance remains sufficient. For regulated domains operating heterogeneous stacks, the gap is structurally elevated and the promotion gate becomes an operational obligation, indissociable from the regulatory obligation.

The predictable objection to this criterion is that it shifts the burden from issuing authorities to user organizations, transforming a public recommendation into local work. The objection is just, and that is precisely the point. The burden was never elsewhere. The illusion that it was came from the silence on variants, silence that the implicit doctrine interpreted as coverage. The promotion criterion does not shift the burden, it makes visible the burden that was already there.

Conclusion

Four inoperant public mitigations, one validated empirical mitigation, one operational concept extracted. The sequence is not accidental. It reproduces, on an external terrain and under temporal constraint, what Article V posed in theory for training phases. The promotion gate between artifact and normative dependency is a dispositive that industry cannot externalize. It has no structural owner. It is the responsibility of the organization that operates the target, because it is the organization that bears the consequences. This distribution is not a defect of the public cybersecurity ecosystem. It is its operating condition at scale.

For regulated domains, where a mismanaged promotion gap becomes a regulatory risk before becoming a technical risk, this condition is not optional. It is the operational expression of what governance asks of security. A public recommendation is an artifact until proven to hold on the exact target. The proof does not fall on the issuer. It falls on whoever operates.

Twingital Institute. Doctrinal series on AI governance architecture, Article VI. The author, Jérôme Vetillard, is VP R&D & Engineering and Chief Product Officer at Qualees (Twingital Ventures). The Twingital Institute publishes architectural and epistemic positions on AI and digital twin systems applied to regulated domains.