Admissibility of Simultaneous Claims in Verification and Validation Workflows

by PanXnubis Gaia Ladrieh
1/3/26

Abstract

Verification and Validation (V&V) methodologies are routinely applied to complex systems under assumptions that are treated as admissible by default. These assumptions – often expressed as guarantees, independence conditions, or bounding arguments – are typically evaluated through testing, simulation, or probabilistic analysis. This paper argues that such approaches implicitly presuppose a prior structural condition: that the set of claims being evaluated can, in principle, all be true at once. We introduce a formal notion of *admissibility* defined as the simultaneous satisfiability of asserted claims, independent of empirical verification. We show that several widely used V&V methodologies rely on claim sets that fail this admissibility condition, not due to incorrect modeling or insufficient testing, but due to internal structural incompatibilities. These failures occur upstream of verification, validation, and uncertainty quantification, and therefore cannot be resolved by additional data, simulation fidelity, or coverage expansion.

1. Introduction

Verification and Validation frameworks are designed to establish confidence that a system behaves as intended under specified conditions. Over the past several decades, increasingly sophisticated techniques have been developed to address uncertainty, complexity, and rare-event behavior, including probabilistic uncertainty quantification, redundancy-based fault tolerance, and large-scale model validation pipelines.

These techniques are typically evaluated on operational criteria: accuracy, conservativeness, coverage, or robustness. Less frequently examined is a more fundamental question: whether the *set of claims* a method relies upon is internally coherent. In practice, V&V workflows often proceed as though this coherence were guaranteed, treating verification as the first line of defense against incorrect or unsafe conclusions.

This paper examines a different precondition. Before a system can be verified or validated, the claims it asserts must be *admissible* – that is, there must exist at least one coherent model of the world in which all asserted properties hold simultaneously. If no such model exists, then verification becomes ill-posed: no amount of testing can confirm a configuration that cannot, even in principle, be realized.

Scope clarification. This work makes no claim regarding the frequency or likelihood of inconsistent assumptions in practice. It addresses only the formal admissibility of simultaneously asserted constraints and the consequences of permitting such assertions to enter verification pipelines without prior evaluation.

The contribution of this work is not a critique of specific techniques, nor a proposal to replace existing V&V practices. Instead, it introduces a minimal structural check that precedes verification: an admissibility evaluation of claim sets. We demonstrate that several commonly accepted methodologies

fail this check in systematic and repeatable ways, independent of implementation quality or data availability.

2. Scope and Non-Goals

The scope of this paper is deliberately narrow. We focus exclusively on the *structural admissibility* of claim sets used in V&V contexts.

Specifically, this work does **not:**

- Assess the correctness of any particular model or system
- Evaluate empirical performance, safety margins, or operational outcomes
- Propose alternative verification or validation methodologies
- Argue against the practical utility of existing techniques

Likewise, this work does **not** claim that systems employing the examined methodologies are unsafe, incorrect, or invalid in practice. Many such systems demonstrably function as intended.

Lifecycle placement. The admissibility evaluation described here is positioned strictly upstream of verification and validation activities. Its function is to determine whether the set of explicit assumptions admitted into downstream analysis can be jointly true at all. Verification results obtained from inadmissible assumption sets are treated as formally uninterpretable, irrespective of numerical accuracy or statistical confidence.

The sole question addressed is whether the claims commonly asserted within these methodologies can all be simultaneously satisfied within a single coherent model of the system and its environment. Where this condition fails, the resulting issue is not one of insufficient verification, but of inadmissible formulation.

3. Admissibility as Pre-Verification Condition

3.1 Definition of Admissibility

Let a system or methodology assert a finite set of claims:

$C = \{c\_1, c\_2, \ldots, c\_n\}$

Each claim represents an explicit statement about system behavior, structure, uncertainty, or guarantees. We define the claim set C to be *admissible* if and only if there exists at least one coherent model M such that:

$M \models \bigwedge_{i=1}^{n} c\_i$

Definition (Admissibility). A set of assumptions is admissible if and only if all stated constraints can be simultaneously satisfied without semantic reinterpretation or probabilistic relaxation.

That is, all claims are simultaneously true within the same model.

This definition is intentionally minimal. It does not require that the model be likely, optimal, or representative – only that it be logically and structurally possible.

3.2 Relationship to Verification and Validation

Admissibility is not a verification activity. It does not involve testing, simulation, or empirical comparison. Instead, it serves as a necessary precondition for those activities.

If a claim set is inadmissible – if no model exists in which all claims hold simultaneously – then verification efforts are necessarily ambiguous. Positive test results cannot establish global correctness, and negative results cannot be meaningfully interpreted, because the underlying formulation lacks a realizable reference state.

In this sense, admissibility is upstream of both verification and validation. It determines whether the questions posed by a V&V workflow are well-formed prior to any attempt to answer them.

4. Methodological Implications

Admissibility failures differ from modeling errors in an important way. They do not arise from incorrect parameter values, insufficient fidelity, or unmodeled dynamics. Instead, they arise from the joint assertion of claims whose meanings, when taken together, exceed what any single system configuration can satisfy.

Such failures are often obscured in practice because individual claims appear reasonable when considered in isolation. It is only when they are required to hold simultaneously – often implicitly – that contradictions emerge.

In the following sections, we examine four widely used methodological contexts in verification and validation. In each case, we explicitly enumerate the claims typically asserted and evaluate their admissibility under the definition given above. Each analysis concludes with a concise structural summary reflecting the outcome of admissibility evaluation.

On metrics. Because admissibility governs whether verification outputs are meaningful at all, comparative performance metrics, error reductions, or predictive improvements are intentionally out of scope. The admissibility gate either holds or fails; no intermediate optimization is implied.

5. Probabilistic Uncertainty Quantification

Probabilistic uncertainty quantification (UQ) methods are widely used to assess the impact of uncertain inputs on system behavior. In verification and validation workflows, these methods are commonly employed to propagate uncertainty, estimate reliability, and support safety or margin-related conclusions. Monte Carlo sampling, in particular, has become a standard tool due to its generality and scalability.

Typical probabilistic UQ formulations rely on a small number of foundational assumptions. Input quantities are modeled as random variables with specified probability distributions. Dependencies between inputs are either explicitly modeled or assumed to be negligible. System behavior is then evaluated across sampled realizations, producing distributions over outputs of interest. From these results, conclusions are drawn regarding expected performance, variability, and extreme outcomes.

Within V&V contexts, probabilistic UQ is often invoked to support claims about worst-case behavior. Tail statistics, confidence intervals, or exceedance probabilities are interpreted as bounding statements on system risk. These interpretations are frequently described as conservative, particularly when independence assumptions or safety factors are employed.

However, when the underlying claims are examined simultaneously, a structural tension emerges. On the one hand, probabilistic UQ asserts that input uncertainties are meaningfully represented by probability distributions, often calibrated from data or expert judgment. On the other hand, worst-case or bounding interpretations require that rare or extreme realizations be treated as representative of system-limiting behavior.

These two claims operate at different structural levels. Probabilistic modeling treats tail events as unlikely by construction, while worst-case reasoning treats those same events as determinative. Reconciling these perspectives requires additional assumptions about coupling, simultaneity, or physical realizability that are rarely stated explicitly and are not identifieable from the available evidence.

Moreover, the identifiability of the assumed input distributions themselves is limited. Distinct distributional forms – particularly in the tails – can produce indistinguishable results over finite samples. As a result, claims about the adequacy of a particular probabilistic representation cannot be uniquely supported by observable evidence alone.

The issue identified here is not one of insufficient sampling, poor calibration, or inappropriate use of Monte Carlo methods. Rather, it arises from the simultaneous assertion of probabilistic independence, tail-based bounding behavior, and worst-case interpretability within a single formulation.

Under the stated assumptions, the claim set is structurally inadmissible: tail-based worst-case assertions and distributional independence cannot be simultaneously satisfied.

6. AI Model Validation

Validation of machine learning and artificial intelligence models presents a distinct but related set of challenges in V&V workflows. Such models are increasingly deployed in contexts where reliability, robustness, and safety are critical, prompting the development of extensive validation pipelines based on testing, benchmarking, and performance evaluation.

Standard AI validation practices assert several core claims. Model performance on held-out test data is taken as evidence of generalization. Validation datasets are assumed to be representative of the operational environment. Coverage metrics, stress testing, and scenario exploration are used to probe model behavior under varied conditions. Together, these practices are used to justify claims about acceptable performance within a defined scope.

At the same time, it is widely acknowledged that complex models may exhibit unanticipated failure modes, particularly when exposed to novel inputs, distributional shifts, or rare combinations of features. This acknowledgment motivates ongoing monitoring, retraining, and post-deployment safeguards.

When considered independently, each of these claims is reasonable. However, when asserted

simultaneously, a structural incompatibility appears. Representativeness implies that the validation dataset captures the relevant features and behaviors of the operational domain. The acknowledgment of unknown or unenumerated failure modes implies the opposite: that relevant behaviors may exist outside the observed data.

No amount of test coverage can resolve this tension. Expanding datasets, improving benchmarks, or increasing evaluation rigor does not alter the underlying structure of the claims. Two models with materially different failure surfaces may produce indistinguishable validation metrics, rendering generalization guarantees non-identifiable from observed performance alone.

The result is not a failure of AI validation as a practice, but a limitation of the claims that can be supported by validation evidence. Assertions of safety, reliability, or completeness depend on properties that cannot be uniquely inferred from finite observations, regardless of scale or sophistication.

Under the stated assumptions, the claim set is structurally inadmissible: representativeness and the existence of unknown failure modes cannot be simultaneously satisfied.

7. Fault-Tolerant Systems and Redundancy

Fault-tolerant system design relies on redundancy to mitigate the impact of component failures. In safety-critical and high-reliability domains, redundancy is employed to ensure continued operation in the presence of faults, often under the assumption that individual component failures are rare and statistically independent.

Standard formulations assert several core claims. Redundant components are assumed to fail independently, such that the probability of simultaneous failure is substantially reduced. System architectures are designed to tolerate a defined number of faults, and verification activities focus on demonstrating correct behavior under specified fault scenarios. Together, these claims support broader assertions regarding system-level reliability and safety.

In practice, fault tolerance analyses often combine probabilistic reasoning with worst-case guarantees. Independence assumptions are used to justify low joint failure probabilities, while architectural arguments are used to assert that even adverse fault realizations are manageable. These perspectives are typically treated as complimentary.

When examined simultaneously, however, a structural incompatibility arises. Independence assumptions characterize failures as uncorrelated and non-simultaneous, while worst-case system-level guarantees require the system to remain safe under coordinated or simultaneous fault realizations. The latter implicitly assumes a degree of coupling that the former explicitly denies.

The incompatibility is not resolved by introducing additional redundancy or refining failure models. Empirical evidence of independence is inherently limited, as correlated failure modes may remain unobserved until triggered by specific environmental or operational conditions. As a result, the independence of failures cannot be uniquely identified from system behavior prior to failure.

The admissibility issue identified here does not imply that fault-tolerant systems are ineffective. Rather,

it reflects a structural limitation in the simultaneous assertion of probabilistic independence and worst-case system guarantees within a single claim set.

Under the stated assumptions, the claim set is structurally inadmissible: independence of failures and worst-case system-level guarantees cannot be simultaneously satisfied.

8. Diverse Redundancy

Diverse redundancy extends traditional fault-tolerant design by employing multiple, independently developed implementations of the same function. This approach is commonly sued in contexts where common-mode failures are a concern, with the goal of reducing the likelihood that a single defect or design flaw compromises the entire system.

Claims associated with diverse redundancy typically include the assertion that implementation diversity mitigates shared defects, that independent development processes reduce correlated failures, and that the resulting system exhibits improve reliability. These claims are often used to justify strong safety or assurance statements, particularly in environments where single-implementation risk is considered unacceptable.

Unlike simple redundancy, diversity-based approaches introduce additional complexity in how failure modes are characterized and bounded. Diversity is often treated as a qualitative property inferred from differences in design, language, tooling, or development teams. Residual risk is then assumed to be acceptably low based on the absence of observed common-mode failures.

A structural tension emerges when these claims are examined jointly. Meaningful diversity implies that implementations do not share failure modes in any systematic way. Bounding residual risk, however, requires an assumption that the space of possible failures is sufficiently characterized and constrained. These two requirements pull in opposite directions: the more unconstrained the failure space, the less defensible a global risk bound becomes.

Furthermore, there exists no operationally identifiable metric that distinguishes true semantic diversity from superficial variation. Distinct implementations may behave identically under untested conditions, and no finite validation effort can establish the absence of shared failure modes. As a result, claims about bounded residual risk depend on properties that cannot be uniquely inferred from observable behavior.

The admissibility failure in this context arises not from flawed engineering practice, but from the simultaneous assertion of diversity and boundedness within a single system-level claim set.

Under the stated assumptions, the claim set is structurally inadmissible: meaningful implementation diversity and globally bounded residual risk cannot be simultaneously satisfied.

9. Implications for Verification and Validation Practice

The analyses presented in Sections 5 through 8 identify a class of failures that are distinct from modeling error, implementation defect, or insufficient testing. In each case, the failure arises from the simultaneous assertion of claims that cannot all be satisfied within a single coherent model of system and its environment. These failures therefore occur upstream of verification and validation activities.

From a practical standpoint, this distinction matters. Verification and validation methods are designed to evaluate whether a system meets specified requirements under defined assumptions. When the assumptions themselves are inadmissible, verification outcomes become inherently ambiguous. Positive results cannot establish global correctness, while negative results cannot be conclusively attributed to implementation deficiencies rather than structural inconsistency.

Introducing an explicit admissibility evaluation prior to verification does not require changes to existing V&V workflows. Rather, it provides a structural filter that clarifies the logical status of the claims being evaluated. By identifying inadmissible claim sets early, practitioners can avoid downstream effort spent validating formulations that cannot, even in principle, be jointly satisfied.

Importantly, admissibility evaluation does not render probabilistic analysis, redundancy, or model validation obsolete. These methods remain valuable within their appropriate scope. The contribution of admissibility analysis is to delineate that scope explicitly, ensuring that claims derived from these methods do not exceed what their underlying structure can support.

10. Conclusion

This paper introduced a formal notion of admissibility as a pre-verification condition for verification and validation workflows. Admissibility was defined as the existence of at least one coherent model in which all asserted claims are simultaneously true. This criterion is independent of empirical testing, probabilistic reasoning, or model fidelity.

Applying this criterion to four widely used methodological contexts – probabilistic uncertainty quantification, AI model validation, fault-tolerant redundancy, and diverse redundancy – revealed a consistent pattern. In each case, commonly asserted claim sets fail admissibility due to internal structural incompatibilities or reliance on non-identifiable properties.

These findings do not invalidate existing V&V practices, nor do they imply that systems employing these methods are unsafe or incorrect. Instead, they highlight a structural gap upstream of verification: the absence of an explicit check on whether the claims being evaluated can all be true at once.

Incorporating admissibility evaluation into V&V workflows offers a minimal but meaningful refinement. By ensuring that verification operates on well-posed claim sets, it strengthens the interpretability and rigor of downstream validation activities without displacing established methodologies.

Appendix A: Formal Admissibility Evaluation

A.1 Claim Sets and Simultaneous Satisfiability

Let a system, model, or methodology assert a finite set of claims:

$C = \{c_1, c_2, \ldots, c_n\}$

Each claim c_i is an explicit statement about system properties, guarantees, or assumptions. A claim set C is said to be *admissible* if there exists at least one model M such that:

$M \models \bigwedge_{i=1}^{n} c_i$

That is, all claims are simultaneously satisfied within the same model.

If no such model exists, the claim set is *inadmissible*. This determination is structural and does not depend on likelihood, optimality, or empirical frequency.

A.2 Contradiction Detection

A contradiction arises when two or more claims impose mutually exclusive requirements on the same underlying system realization. Importantly, such contradictions may not be apparent when claims are evaluated independently.

The admissibility criterion therefore requires joint evaluation of all claims. Ambiguity is not resolved in favor of coherence; if coherence is asserted, it must be demonstrable by construction.

A.3 Identifiability Constraint

In addition to logical consistency, admissibility requires that claims not dependent on non-identifiable properties.

Let Θ denote the set of possible true system properties and D the set of all observable data obtainable through testing, simulation, or monitoring. A property $\theta \in \Theta$ is identifiable if:

$$\sum \theta\_1 \neq \theta\_2,\ P(D \mid \theta\_1) \neq P(D \mid \theta\_2)$$

If two materially different systems properties yield indistinguishable observations, then claims that depend on distinguishing between them are not admissible.

Claims of completeness, bounded residual risk, or absence of unknown failure modes frequently violate this condition.

A.4 Admissibility Gate (Abstract)

The admissibility evaluation used in this paper can be expressed abstractly as follows:

- Enumerate the explicit claim set C
- Determine whether a model M exists that satisfies all claims simultaneously
- Verify that no claim depends on a non-identifiable property

If either condition fails, the claim set is inadmissible.

A.5 Implementation Note

The analysis presented in this paper were conducted using a general-purpose admissibility evaluation kernel designed to operate independently of domain-specific models, probabilistic assumptions, or empirical data. The kernel evaluates claim sets purely on structural and logical grounds and is intended to be applied upstream of verification and validation activities. No domain-specific behavior is encoded in the kernel itself.