

# Scenario-based safety assessment of AI-enhanced automated driving systems

Verification and validation (V&V) are critical in the automotive industry to ensure vehicle safety, especially with the rise of automated driving systems. These systems introduce new complexities due to their reliance on AI and the complex dynamic nature of their operational design domain (ODD). Current V&V approaches struggle with AI's statistical, black-box nature, making it difficult to ensure systems meet their requirements in all situations.

Our work identifies several challenges in V&V of AI-based systems:

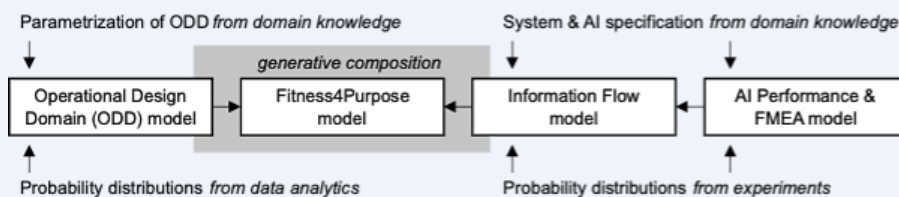
- **Environment:** Specifying the ODD mathematically is complex, and current standards like ISO 21448 and UL 4600 provide guidelines but are insufficient for data-driven AI.
- **Requirements:** Safety and predictable behavior are key requirements, but specifying correct behavior in every situation is difficult due to the complexity of the environment.
- **System Understanding:** AI systems are hard to interpret because of their non-linear, non-transparent models.
- **Testing:** proving the safety of AI system through testing is difficult due to the vast input feature domain and the need to quantify risks based on finite tests.



**Figure 1 - Semantic gap:** the missing connection of symbolic concepts and sub-symbolic pixels. The icons on the left depict all manners in which pedestrians might occur on the symbolic level. Depending on the conditions like weather and light, the right side of the figure depicts how this is translated to pixels in a camera on the sub-symbolic level.

TNO proposes a scenario-based V&V methodology for systems containing AI functionality, based on quality of information exchanged between components:

- **Parameterized ODD Description:** Creating a probabilistic model of the ODD based on scenarios. Scenarios provide a widely accepted way to describe the real-world driving environment. The probabilistic model helps in understanding the conditions under which the system operates and the potential risks involved.
- **FMEA of AI:** Establishing causal relationships between AI failures and their consequences to assess risk impact. This helps in identifying potential failure modes and their effects on system performance.
- **Information Flow Model:** Understanding the quality of the information flow within the system to assess its fitness for purpose. This model integrates the probabilistic ODD description and FMEA to evaluate system performance in various scenarios.



**Figure 2:** Fitness-for-purpose model originating from composition from other models.

This methodology helps bridge the gap between symbolic requirements and sub-symbolic AI components and provides a way to minimize the number of unanalyzed, unsafe scenarios while keeping the testing load manageable.

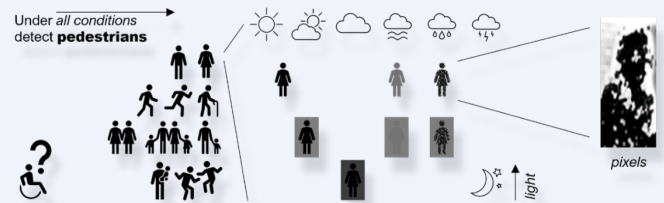
Source: Paardekooper J.P., Borth M., "Toward a Methodology for the Verification and Validation of AI-Based Systems", 2025 SAE International

# AI強化自動運転システムのシナリオベース安全性評価

自動車業界では、特に自動運転システムの台頭に伴い、車両の安全性を確保するための検証と妥当性確認（V&V）が重要です。これらのシステムは、AIに依存し、その運用設計領域（ODD）の複雑な動的性質により、新たな複雑さをもたらします。現在のV&Vアプローチは、AIの統計的なブラックボックス性により、すべての状況でシステムが要件を満たすことを保証することは難しいです。

私たちの研究は、AIベースシステムのV&Vにおけるいくつかの課題を特定しています：

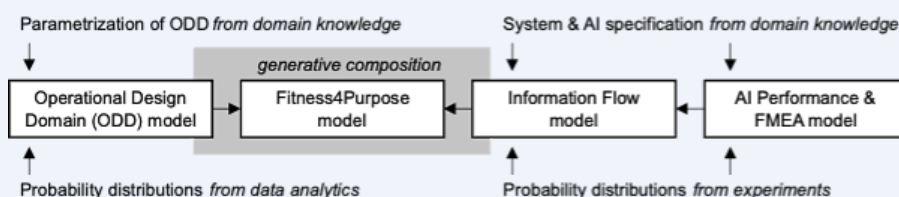
- ・**環境**：ODDを数学的に指定することは複雑であり、ISO 21448やUL 4600のような現在の標準は、データ駆動型AIには不十分です。
- ・**要件**：安全性と予測可能な行動が重要な要件ですが、環境の複雑さのためにすべての状況で正しい行動を指定することは困難です。
- ・**システム理解**：AIシステムは、その非線形で不透明なモデルのために解釈が難しいです。
- ・**テスト**：広範な入力特徴ドメインと有限のテストに基づいてリスクを定量化する必要があるため、AIシステムの安全性をテストで証明することは困難です。



**Figure 1 - Semantic gap:** the missing connection of symbolic concepts and sub-symbolic pixels. The icons on the left depict all manners in which pedestrians might occur on the symbolic level. Depending on the conditions like weather and light, the right side of the figure depicts how this is translated to pixels in a camera on the sub-symbolic level.

TNOは、AI機能を含むシステムのためのシナリオベースのV&V方法論を提案しています。これは、コンポーネント間で交換される情報の品質に基づいています：

- ・**パラメータ化されたODD記述**：シナリオに基づいてODDの確率モデルを作成します。シナリオは、現実の運転環境を記述するための広く受け入れられた方法を提供します。確率モデルは、システムが動作する条件と関係するリスクを理解するのに役立ちます。
- ・**AIのFMEA**：AIの故障とその結果との因果関係を確立し、リスク影響を評価します。これにより、システム性能に対する潜在的な故障モードとその影響を特定するのに役立ちます。
- ・**情報フローモデル**：システム内の情報フローの品質を理解し、その目的適合性を評価します。このモデルは、確率的ODD記述とFMEAを統合し、さまざまなシナリオでのシステム性能を評価します。



**Figure 2:** Fitness-for-purpose model originating from composition from other models.

この方法論は、象徴的な要件とサブ象徴的なAIコンポーネントの間のギャップを埋めるのに役立ち、テスト負荷を管理可能に保ちながら、未分析の安全でないシナリオの数を最小限に抑える方法を提供します。

出典: Paardekooper J.P., Borth M., "Toward a Methodology for the Verification and Validation of AI-Based Systems", 2025 SAE International