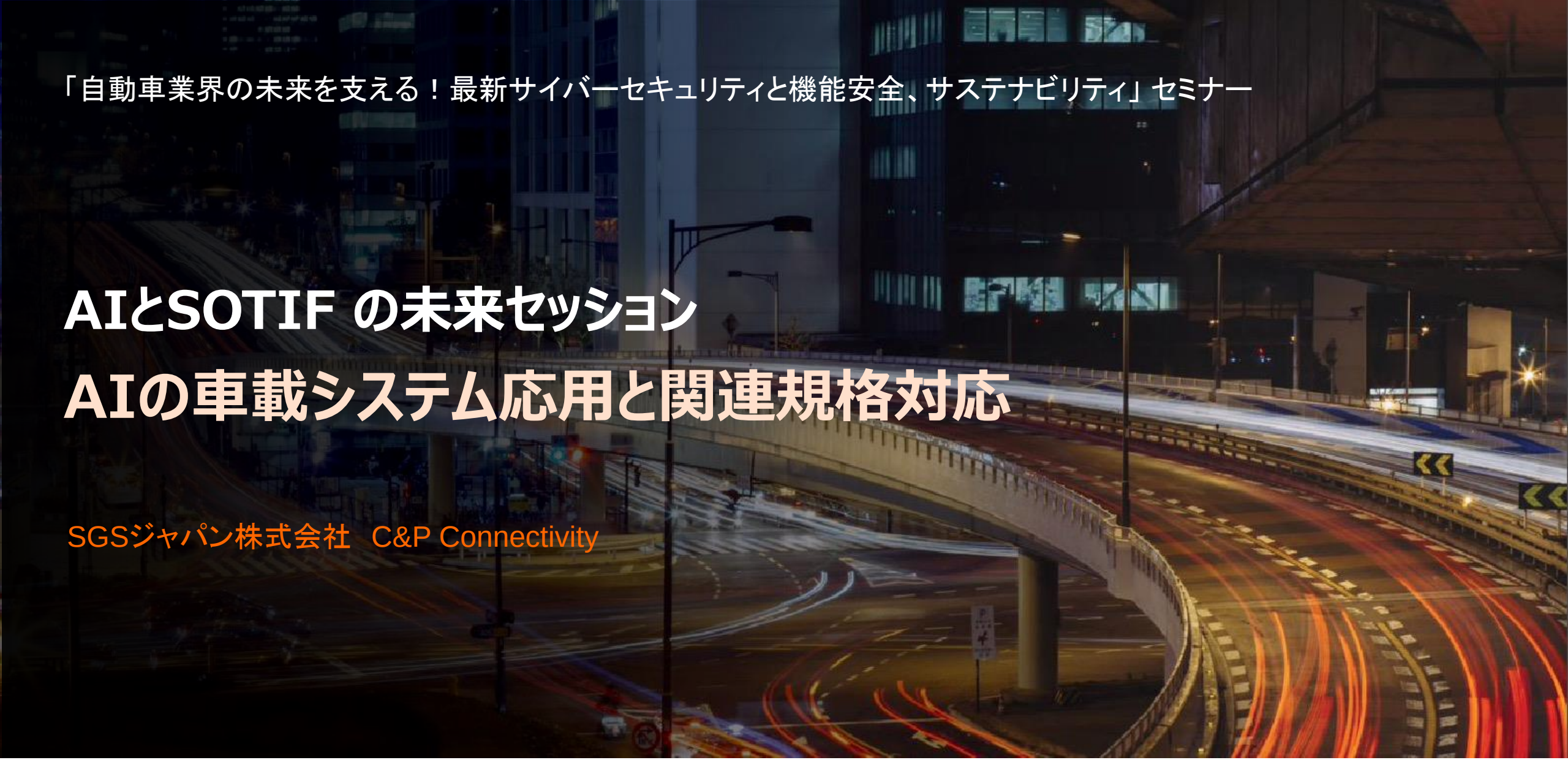


「自動車業界の未来を支える！最新サイバーセキュリティと機能安全、サステナビリティ」セミナー

AIとSOTIF の未来セッション
AIの車載システム応用と関連規格対応

SGSジャパン株式会社　C&P Connectivity

2025年5月23日

自己紹介

岡本 謙一
SGSジャパン株式会社
C&P Connectivity 機能安全
kenichi.okamoto@sgs.com

- 略歴
  - 車載部品Tier2
    - 電動パワーステアリングECUのセーフティマネジメント、他
  - 大型車OEM
    - 自動運転(Lv4)先行開発マネジメント、他
- 2023年よりSGSジャパンにてSOTIF、機能安全業務に従事
  - *SGS-TÜV認定の有資格者 (**AFSE, IFSE, AISE**, SOTIF PRO, SC-AFSP, CACSP)*
  - *Automotive SPICE provisional assessor*
  - *Deep Learning for GENERAL : JDLA Certificate*

AFSE IFSE AISE

2020 #2

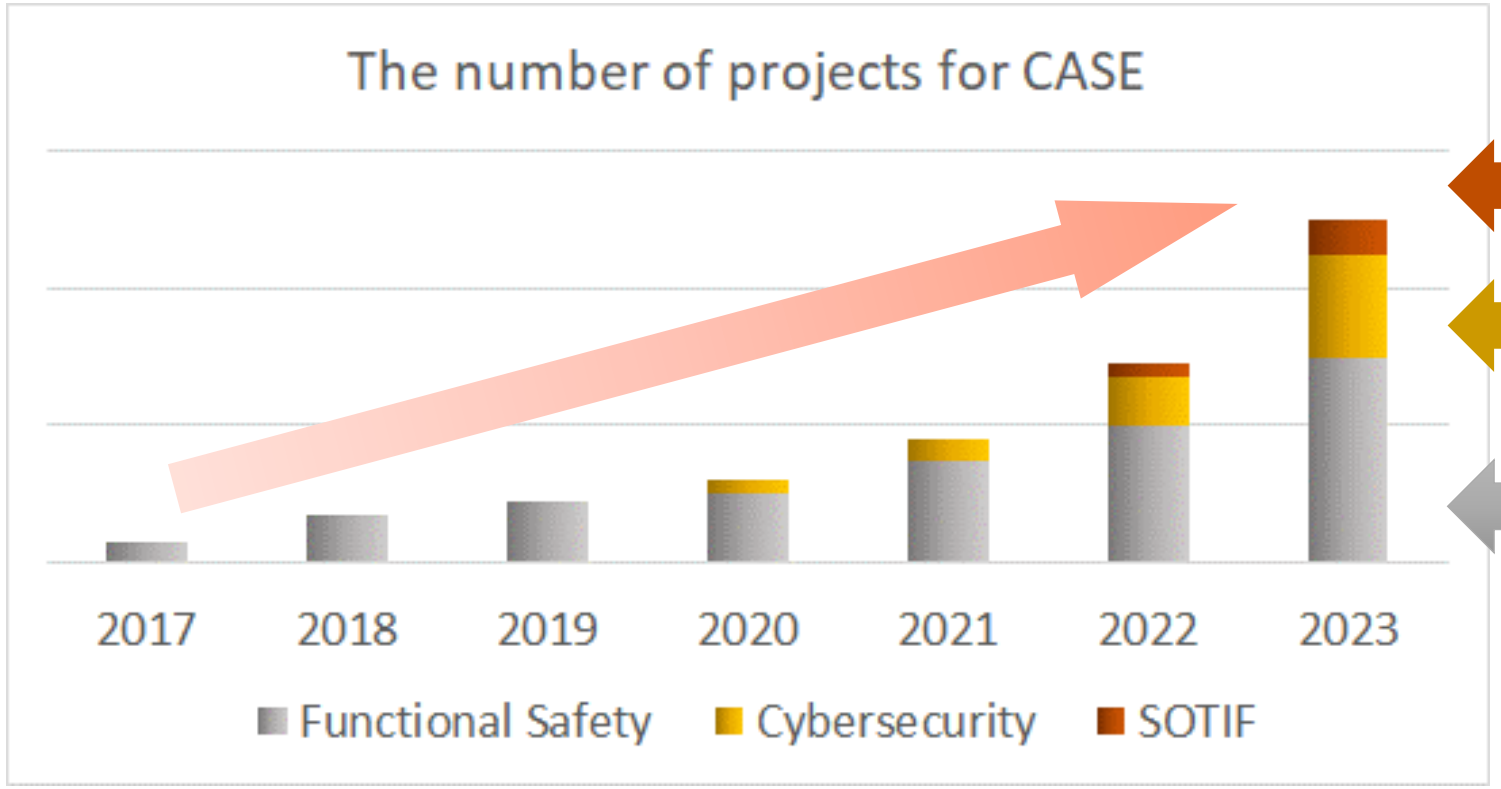
アジェンダ

- 法規、規格、標準の制定動向
- AI関連規格：A-SPICE(MLE)、ISO/PAS 8800
- AI安全要件の導出
- AI安全分析
- AI安全性論証
- Safety Management Systemの備え
- まとめ

法規、規格、標準の制定動向

分類	法規、規格、標準	動向
品質管理	• ASPICE	• Automotive SPICE 4.0の対応に向けたプロセス構築 • 機械学習（MLE）対応開始
機能安全規格	• ISO 26262 • ISO PAS 8800	• ISO26262 3rd Edition 議論開始 (Leader of WG) - Automated driving (GM) - Connected vehicle (Continental) - New Energy Vehicle (CATARC) - Safety demonstration for AI/DL (NVIDIA) etc… • 2024年12月にISO PAS 8800 AIの安全性に関する規格発行
自動運転法規／規格	• UN-R157 (ALKS) • UN-R171(DCAS) • ISO 21448:2022 • ISO 34502:2022 • ISO 34503:2023	• 2024年9月にDCAS（自動運転Lv2)発効、ハンズオフ要件が継続検討中のため日本では任意適用 • 意図機能の安全性に関する規格 ISO 21448 (SOTIF)が発行 • シナリオベースの安全性評価、ODD定義、シナリオ定義に関する規格発行
セキュリティ法規／規格	• UN-R155/156 • ISO/SAE 21434	• EUでは、2025年7月に全ての新車を対象にUN-R155とUN-R156の義務化 • 開発製品におけるセキュリティ対応の必要性を明確化するため、TARAや脆弱性分析を活用したアサンプションを行うケースが増加

SGSジャパンの経験に基づくトレンド



SOTIF対応はOEMとTIer1で徐々に開始されている

自動運転に向けたサイバーセキュリティ対応の強化

AD/ADASとEV関連の機能安全開発が世界中で増加

機能安全、サイバーセキュリティ、SOTIFに関する“**最新技術**”の把握が重要

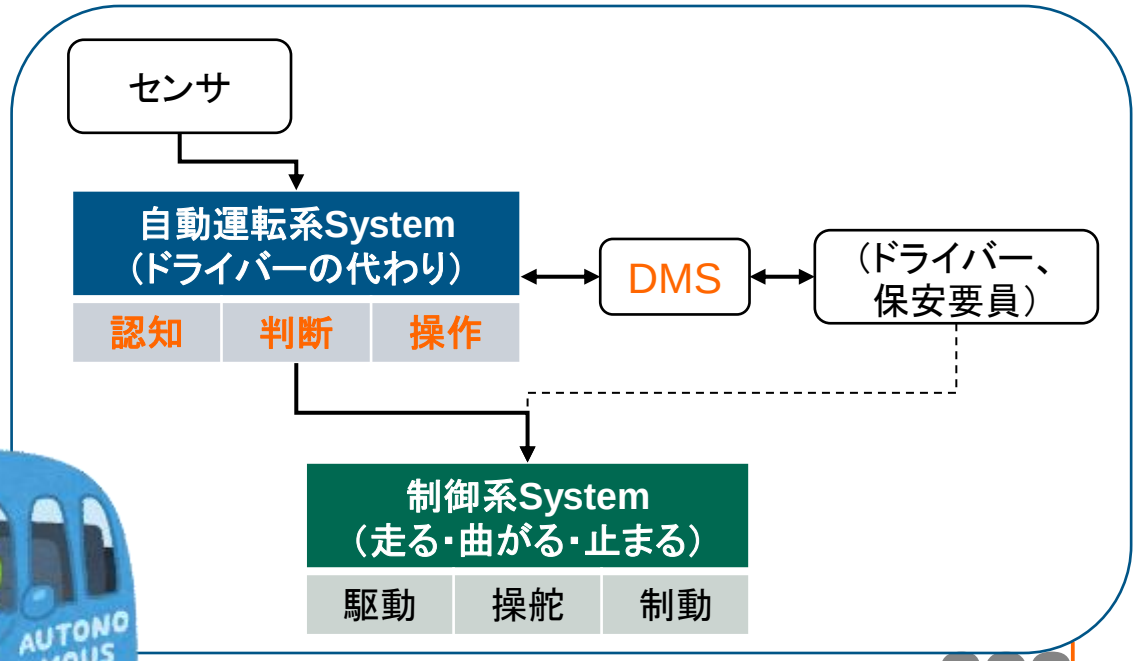
自動車システムの知能化を加速するAI

- **環境認識**: 物体検出・分類（カメラ、LiDAR）、走路認識、標識認識など
- **走行判断・予測**: 周囲車両の軌道予測、自車経路計画、運転行動の予測
- **操作**: 状況に応じた最適なアクセル・ブレーキ・ステアリング制御
- **ドライバーモニタリング (DMS)**: わき見、居眠り検知、ドライバー状態推定。
- **その他**: 車室内インフォテインメント、音声対話システムなど。

A-SPICE(MLE, SUP.11)
（機械学習エンジニアリングプロセス）

- 特に運転支援や自動運転機能において、AIが安全に関わる役割を担うケースに注意

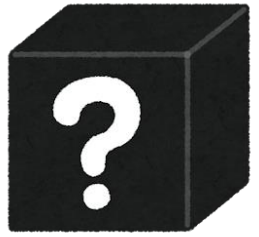
ISO 26262, ISO 21448を
補完し、AIの安全性を論証



AI（機械学習）は何が違うのか？

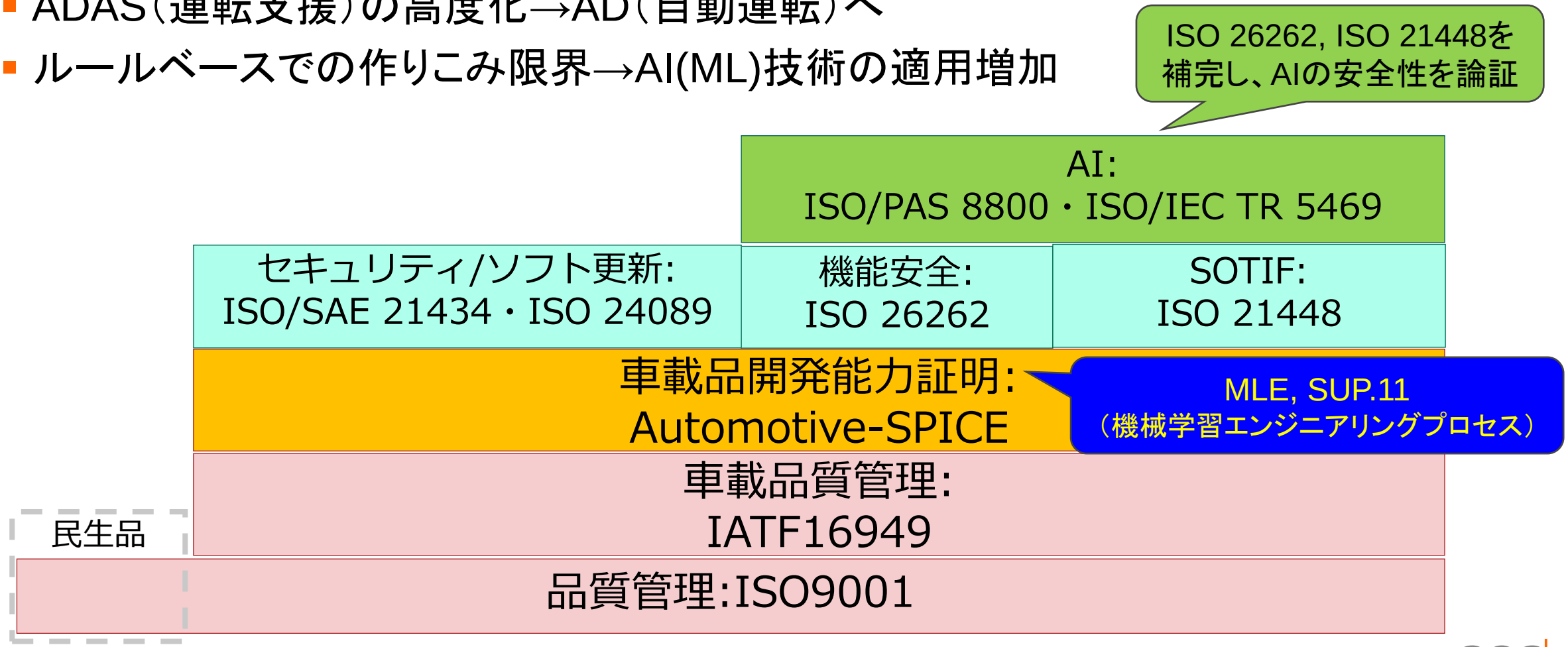
- **データ駆動**: 明示的なロジックではなく、データからパターンやルールを学習する。
- **確率的・非決定的挙動**: 同じ入力でも出力が変動する可能性。常に100%の正解を保証しない。
- **複雑性・不透明性**: 内部の動作原理（なぜその結論に至ったか）を完全に理解・説明することが困難（ブラックボックス性）。
- **データへの高い依存性**: 学習データの量、質、多様性、バイアス（偏り）が性能や挙動に直接影響する。
- **継続的学習**: システム運用後も学習・進化しうる（性能向上と新たなリスクの可能性）。

- これらの特性が、安全性を考える上で新たなチャレンジを生む。



車載規格対応のトレンドと効率化

- ADAS（運転支援）の高度化→AD（自動運転）へ
- ルールベースでの作りこみ限界→AI(ML)技術の適用増加

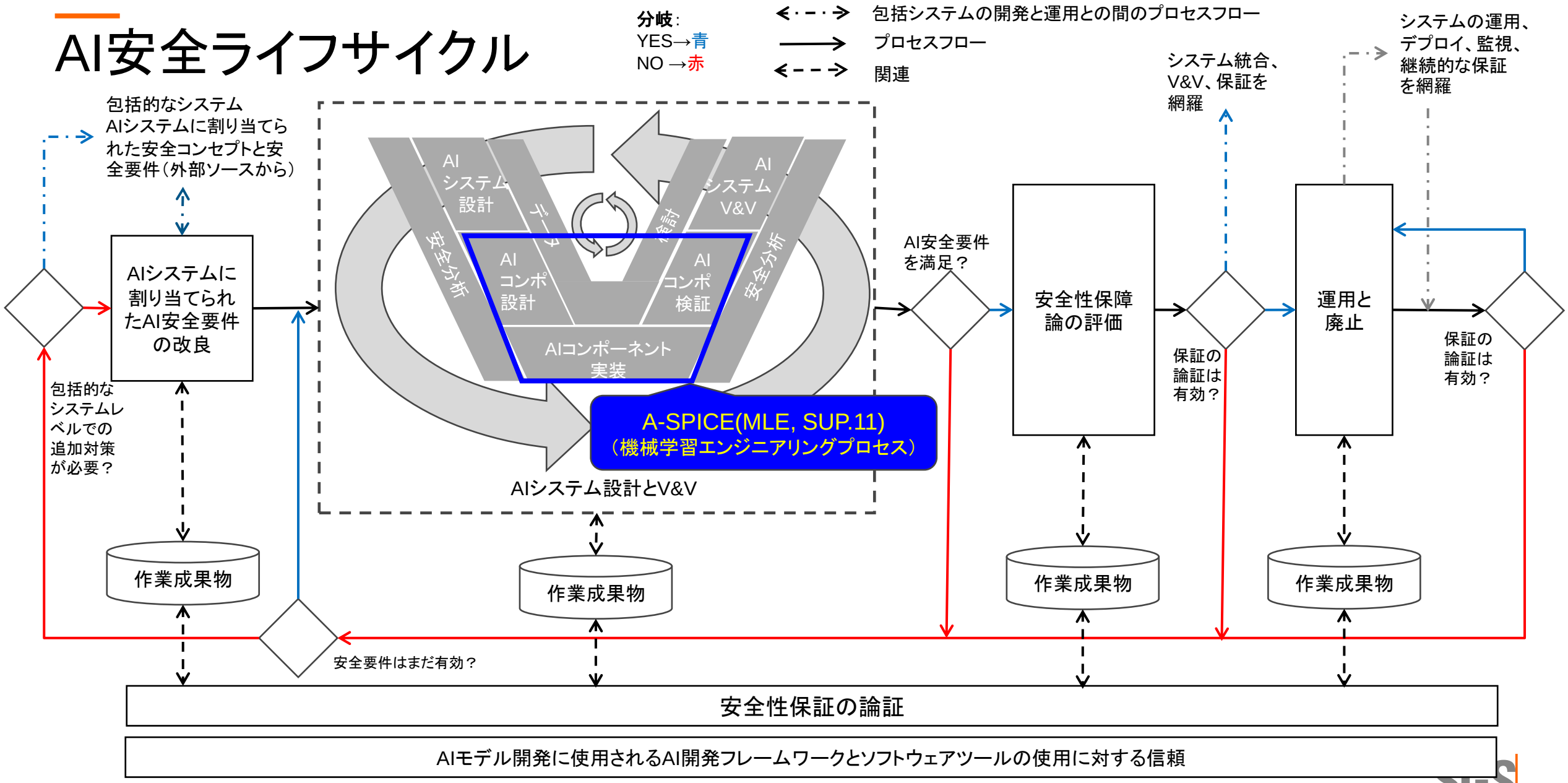


A-SPICE(MLE)とISO/PAS 8800の違い

観点	A-SPICE(MLE)	ISO/PAS 8800
目的	**プロセス能力**の向上、開発リスクの低減	**安全なAIシステム**の実現、安全性の説明責任
主な焦点	MLコンポーネント**開発プロセスの品質**・成熟度	AIシステム全体の安全性の確保と論証
主な課題	プロセスは計画通り適切に実施されているか？	AIシステムは安全目標を達成し、リスクは許容可能か？ 安全性をどう証明するか？
データへの関心	データ管理プロセスは適切か？	データは安全性を保証する上で十分な品質・多様性・網羅性を持つか？バイアスは？
モデルへの関心	モデル開発プロセスは適切か？結果は記録されているか？	モデルは安全要件を満たすか？未知・異常入力に安全に対処できるか？
V&Vへの関心	モデル評価プロセスは適切か？	システム全体の安全性を証明する上でV&V活動は十分か？（特にSOTIF観点）
最終成果物	適切に実施されたプロセスの証拠	安全性の論証（Safety Case）

A-SPICE(MLE)でプロセスを確立し、ISO/PAS 8800でAIシステムの安全性を示す

AI安全ライフサイクル



ISO PAS 8800、図 7-1 – AI の安全性 ライフサイクル

ISO/PAS 8800 の主要な技術的・プロセス的観点

- **AIシステムの特性と安全への影響分析:** AIの特性（確率性、データ依存性等）がどのように安全リスクにつながるか。
- **データ管理:** データ収集、アノテーション、品質確保、完全性、トレーサビリティに関する考慮事項。
- **モデル学習と特性:** モデル選択、学習プロセス、性能評価（汎化性能、ロバスト性）、不確かさ評価。
- **AIコンポーネントの検証・妥当性確認:** AI特有のV&V手法、テスト戦略、メトリクス。
- **システムへの統合:** AIコンポーネントと非AIコンポーネントの相互作用、全体システムとしての安全性確保。
- **AIライフサイクル:** 開発、運用、保守、アップデートにおける安全確保プロセス。

AI安全要件：従来の要件定義との違い

- **従来の要件:** 機能仕様（例：「速度X以上でブレーキを踏んだらY秒以内に停止」）のように、明確な入力と期待される決定的な出力を記述しやすい。

- **AIへの要件:**
- **性能目標:** 「ODD内の条件下で、歩行者検出率99.9%以上、誤検出率0.1%以下」など、確率的・統計的な目標値となることが多い。
- **振る舞いの制約:** 「信頼度が閾値以下の場合は、警告を発しドライバーに制御を促す」など、不確実な状況での安全な振る舞いを規定。
- **ロバスト性要求:** 「特定のノイズレベルや天候条件下でも、性能低下がX%以内であること」。
- **データ要求:** 「学習データは特定のバイアス指標を満たし、Y種類以上のシナリオをカバーすること」。

明示的な機能だけでなく、「どのように振る舞うべきか（性能、ロバスト性、安全性）」を定義する必要がある。

AI安全要件導出のステップ

- システムレベルのハザード分析: FuSa (ISO 26262 HARA) と SOTIF (ISO 21448) の分析を実施し、システムレベルの安全目標・安全要件を定義。
- AI機能への割り当て: システムレベルの安全要件のうち、AIコンポーネントが寄与・担当する部分を特定。
- AI固有ハザードの考慮: AIの特性に起因するハザードを特定・分析。（ISO/PAS 8800の観点を活用）

- AI安全要件の具体化:
  - AIコンポーネントに対する具体的な安全要件を定義（性能、ロバスト性、振る舞い、データ等）。定量化・検証可能な形で記述することが理想。
  - ODDとの関連付け: すべてのAI安全要件は、定義されたODDと明確に関連付けられる必要がある。

ハザード分析例：AIによる道路標識認識システム

- ケース1
  - 「認識結果をドライバーへ提示のみ」
- ハザード例
  - 制限速度を実際より高く誤表示し、ドライバーが速度超過するリスクが増加する。
  - 一時停止標識を表示せず、ドライバーが見落とすリスクが増加する。
- リスク低減：QMレベル

A-SPICE(MLE, SUP.11)
プロセス準拠で対応

- ケース2
  - 「認識結果を使ってブレーキ制御」
- ハザード例
  - 一時停止標識を認識せず、自動ブレーキが作動しない。
  - 道路脇の看板などを一時停止標識と誤認識し、予期せず急ブレーキがかかる。
  - 制限速度を誤認識し、不適切な速度に制御される。
- リスク低減：
  - ISO 26262: ASIL-C～ASIL-D
  - ISO 21448: 受容基準未満

AIへの安全要件を定義して対応

AI安全要件の例とODD

- **要件カテゴリ例:**
- **Accuracy / Performance:** (例) ODD内での物体クラス別検出率、位置推定精度
- **Robustness:** (例) 悪天候、センサーノイズ、一部隠れに対する性能維持、敵対的攻撃耐性
- **Safety / Integrity:** (例) 安全関連判断の信頼度指標、異常検知能力、フェールセーフ挙動
- **Availability:** (例) ODD内での機能提供時間率
- **Data:** (例) 学習データの品質基準、カバレッジ、更新手順

- **ODDの役割:**
- AI機能が安全に動作する前提条件を定義。
- 安全要件が適用される範囲を明確化。
- V&Vのスコープ（どのシナリオでテストすべきか）を決定。

具体的で検証可能なAI安全要件を定義するには、明確なODD定義が不可欠。

課題：性能限界とロバスト性

- **性能限界:**
- **認識・判断の限界**: 予期せぬ物体、複雑な交通状況、悪天候など、学習データでカバーしきれない状況への対応

  →ISO 21448(SOTIF)で対応
- **ODD (Operational Design Domain) の課題:** 設計時に想定した作動領域外での挙動保証の難しさ。

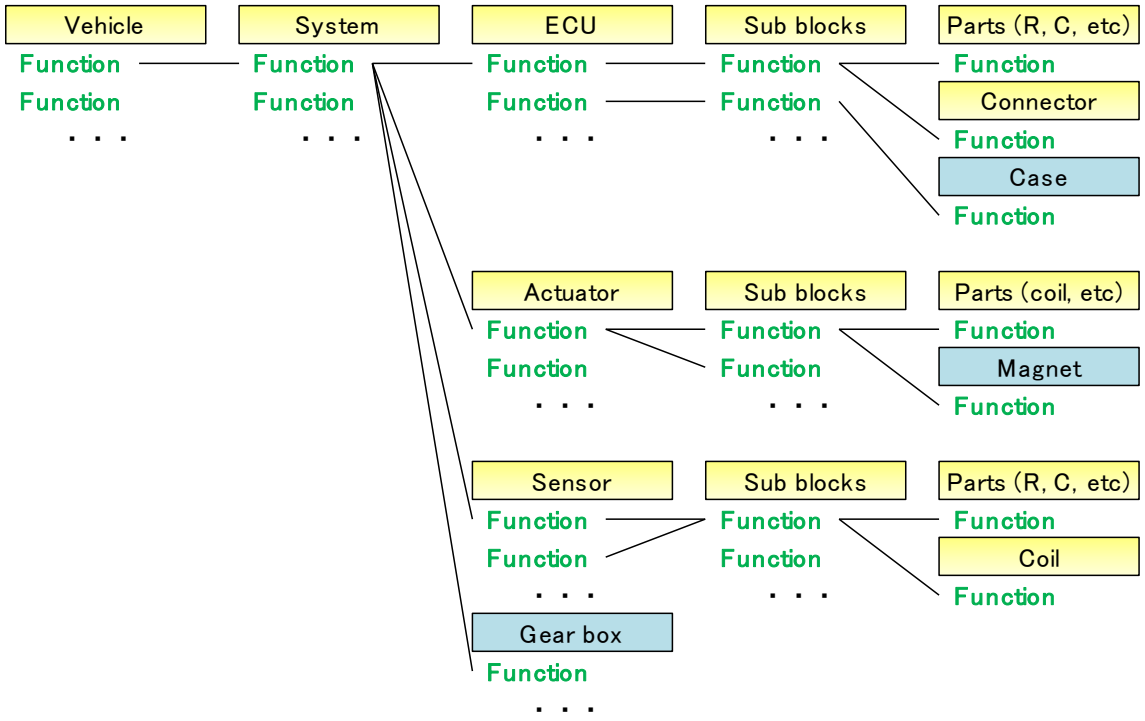
  →ODD境界の検出で対応

- **ロバスト性の問題:**
- **外乱耐性:** センサーノイズ、環境変化（光、天候）に対する性能維持。
- **分布外入力:** 学習データとかけ離れた予期せぬ入力への対応。
- **敵対的攻撃 (Adversarial Attack):** 意図的な微小な改変によるAIの誤動作リスク。

?

AI安全分析①：AI-FMEA / SOTIF-FMEA

- **従来のFMEA:** コンポーネントの故障モード（断線、短絡、SWバグ等）とその影響を分析。



- **AIへの拡張: 故障モードの追加:** AI特有の「故障モード」を考慮する。

  例：「誤分類」「検出漏れ」「誤検出」「出力不安定」「低信頼度出力」「応答遅延」「特定条件下での性能劣化」

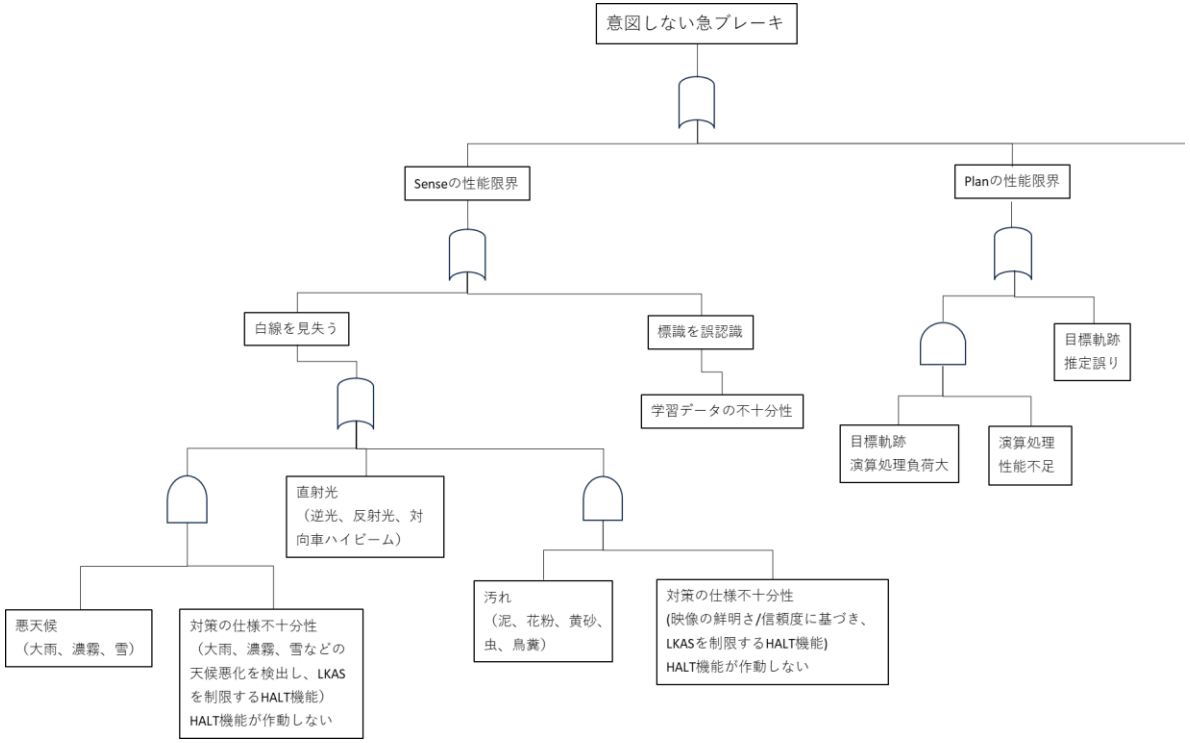
- **原因の深掘り:** これらのAI故障モードがなぜ発生するか？

  例：データ不足、バイアス、アルゴリズム限界、過学習、敵対的入力、センサーノイズ

- **SOTIF観点の統合:** 故障ではないが危険につながる「性能不足」も分析対象とする（SOTIF-FMEA）

AI安全分析②：FTAとシナリオベース分析

- **FTA (Fault Tree Analysis):** トップイベント（望ましくない事象、例：衝突）を設定し、その原因となる基本事象（ハード故障、ソフトエラー、AIの誤動作等）を論理的に展開。AIの誤動作を基本事象として組み込む。
- **シナリオベース分析 (SOTIF/ISO 34502):**
  - 潜在的に危険なシナリオ（例：雨天夜間の高速道路合流）を洗い出す。
  - そのシナリオ下でAIの性能不足や誤動作がハザードにつながるトリガー条件（例：強い雨によるカメラ認識阻害＋レーダーの虚像検出）を特定する。
  - シミュレーションや実データを用いて、発生頻度や影響度を評価。



AI安全分析③：AI固有のリスク探求

- **データ分析:**
  - **カバレッジ分析:** 学習データがODDや想定されるシナリオをどれだけカバーしているか評価。未カバー領域のリスクを特定。
  - **バイアス分析:** データ内の偏りを検出し、それが引き起こす可能性のある不公平・不安全な挙動を分析。
- **モデルロバスト性分析:**
  - **外乱注入テスト:** センサーノイズ等を模擬した入力に対するモデルの挙動を分析。
  - **敵対的攻撃テスト:** モデルの脆弱性を探索し、対策の必要性を評価。

- **説明可能性 (XAI) の活用:** モデルが特定の判断を下した理由を分析し、潜在的な問題（例：間違った特徴量に注目している）を発見。

これらの分析結果は、リスク低減策（モデル改善、データ追加、監視機能など）の検討に繋げる。

検証（Verification） & 妥当性確認（Validation）

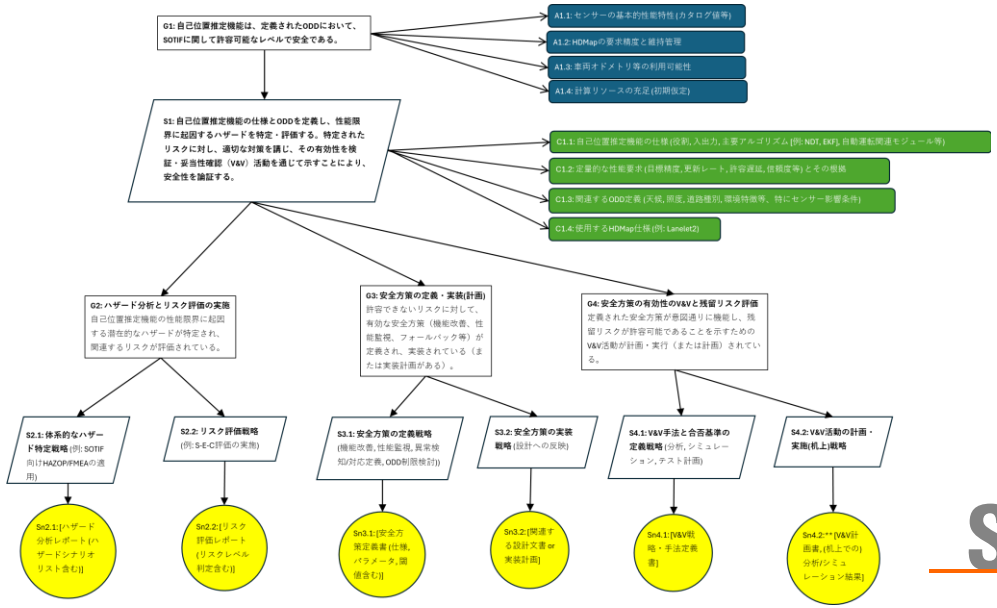
- **目的:** AIコンポーネントが、その仕様（安全要件含む）通りに実装されていることを確認する。
- **主な活動:**
  - **静的解析:** コードレビュー、アーキテクチャレビュー。
  - **単体・結合テスト:** AIモデルを構成するソフトウェアモジュールのテスト。
  - **モデル性能検証:** 定義されたメトリクス（精度、再現率など）に基づき、評価用データセットでモデル性能を定量評価。
  - **ロバスト性検証:** 安全要件で定義された外乱やエッジケースに対する挙動を確認。
  - **データ検証:** 学習・評価データの品質、完全性、トレーサビリティの確認。
  - **要件トレーサビリティ:** 各検証活動がどの安全要件に対応しているか追跡。

- **目的:** AIコンポーネントが、意図された運用環境（システム、ODD）において、安全目標を達成することを確認する。
- **主な活動:**
  - **シミュレーション:** 大量のシナリオを用いた仮想環境でのテスト。コーナーケース、危険シナリオの再現。ツールの忠実度が重要。
  - **ハードウェアインザループ (HIL):** 実機ECUとシミュレーション環境を組み合わせたテスト。
  - **実車テスト:** 制限された環境での段階的なテスト、公道でのフィールドオペレーションテスト(FOT)。カバレッジと安全確保が課題。
  - **シナリオベース評価:** 安全分析で特定された重要シナリオに基づいたテスト。
  - **統計的評価:** 十分な走行距離やシナリオ数に基づき、目標性能を統計的に確認。

AI安全性論証：説得力のある論証のための戦略

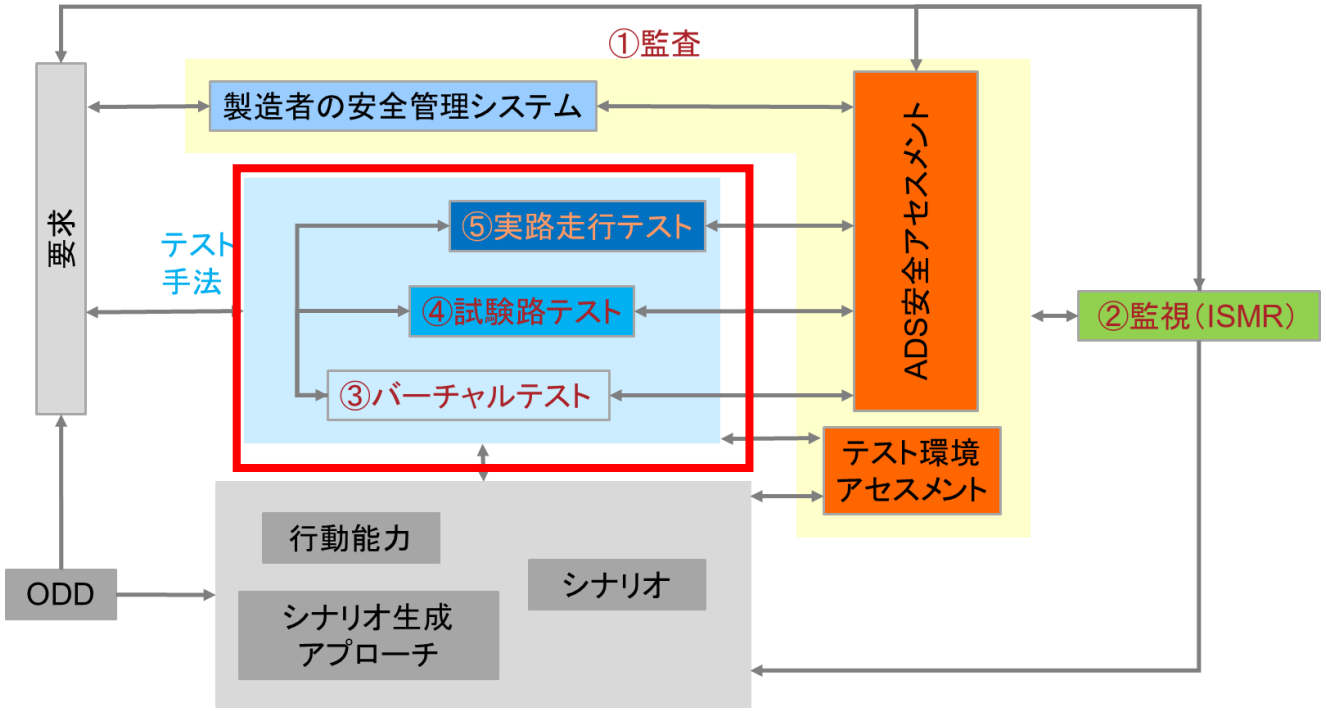
- **安全性論証 (Safety Case):**
  - トップレベルの安全目標（Claim）が達成されていることを示す構造化された議論。
  - 主張（Argument）とそれを裏付ける証拠（Evidence）から構成される。
  - AIに関しては、従来の証拠（テスト結果等）に加えて、データ品質、モデル特性、V&Vの網羅性に関する証拠と論証が重要になる。

1. **多層的なアプローチ (Defense in Depth):** 単一の手段に頼らず、複数の検証・妥当性確認手法、安全機構、モニタリングを組み合わせる。
2. **議論の構造化:** GSN (Goal Structuring Notation) などの表記法を用いて、安全目標から証拠までの論理的な繋がりを明確化。



AI安全性論証：説得力のある論証のための戦略

3. **証拠の多様性**：テストレポート、分析結果、シミュレーションログ、データ品質レポート、プロセス遵守記録、エキスパートレビューなど、多様な証拠を体系的に収集・管理。
4. **定量化と定性評価のバランス**：可能な限り定量的な証拠（性能メトリクス、テストカバレッジ）を提示しつつ、定性的な議論（設計思想、プロセス遵守）で補完。
5. **未知への対応**：AIが学習データに含まれない未知の状況にどう対応するか（安全側に倒れるか等）の論証も重要。



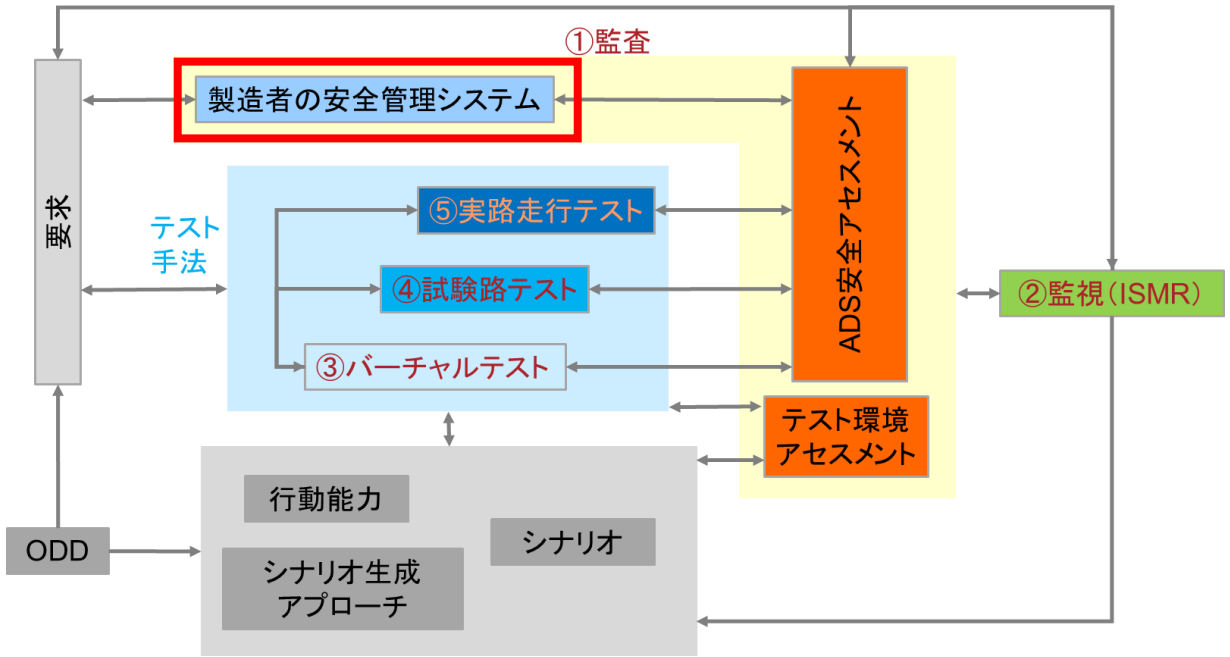
New Assessment/Test Method for Automated Driving (NATM) Guidelines for Validating Automated Driving System (ADS)

Safety Management Systemの備え

- 自動運転Lv3法規：UN-R157(ALKS)
  - Annex4 3.5 Safety Management System
- 自動運転Lv2法規：UN-R171(DCAS)
  - Annex3 3.5 Safety Management System

- SMS の主要コンポーネント: <Annex3-3.5.2>
  - A) 安全方針と目標　Safety Policy and Objectives (SPO):
  - B) 安全リスク管理 Safety Risk Management (SRM):
  - C) 安全保証 Safety Assurance (SA):
  - D) 安全推進 Safety Promotion (SP):

- 特に複雑で新しい技術であるAIの安全性を担保するには、市場に出た後も継続的に安全性を維持・向上させる必要があり、組織的な管理体制（SMS）が求められる。



New Assessment/Test Method for Automated Driving (NATM) Guidelines for Validating Automated Driving System (ADS)

AI固有の考慮事項

- **力量管理**: AI/MLエンジニア、データサイエンティスト、ドメインエキスパート、安全技術者の連携と、それぞれに必要な安全関連スキルの定義。
- **ツール管理**: MLフレームワーク、シミュレータ、データ管理ツール等の選定・利用に関するプロセス（ツール信頼性レベルの考慮）。
- **構成管理**: データセット、学習済みモデル、ソフトウェア、パラメータ等の構成管理とトレーサビリティ確保。

- **プロセス拡張**:
  - **データライフサイクル管理**: データ収集・アノテーション・品質管理・更新に関するプロセス。
  - **モデル開発・学習プロセス**: モデル選択基準、学習環境、ハイパーパラメータ調整、性能評価の標準化。
  - **AI特有のV&Vプロセス**: アシュアランス活動のプロセス定義。
  - **MLOpsと安全プロセス連携**: CI/CDパイプラインと安全活動の統合。

まとめ

- 車載システムの複雑化、知能化に伴い、AI(ML)技術の適用範囲が広がっている

1. 安全性に関わらない部分はA-SPICE(MLE)**のプロセス**で開発
2. 安全性に関わる部分はISO 26262(機能安全)、ISO 21448(SOTIF)に加えてISO PAS 8800の考慮も必要

- 特に複雑で新しい技術であるAIの安全性を担保するには、意図機能が安全(SOTIF)であることの論証が必要。

- 国連(WP29)のGRVAで策定されたNATM（安全性評価フレームワーク）が具体化され、法規適用されている。
  - ALKS(UN-R157) LV3自動運転 ’21年
  - DCAS(UN-R171) Lv2自動運転 ’24年

- 機能安全、SOTIF、サイバーセキュリティを横断的に対応できる組織的な管理体制（SMS）が求められる。

SGSジャパン：規格適用サービス

- Automotive SPICE
- 機能安全(ISO 26262)
- サイバーセキュリティ(ISO 21434)
- SOTIF(ISO 21448)
- AI (ISO/PAS 8800)
- Safety Management System

- SGSでは、6つの段階を踏んで規格適用を支援しています。

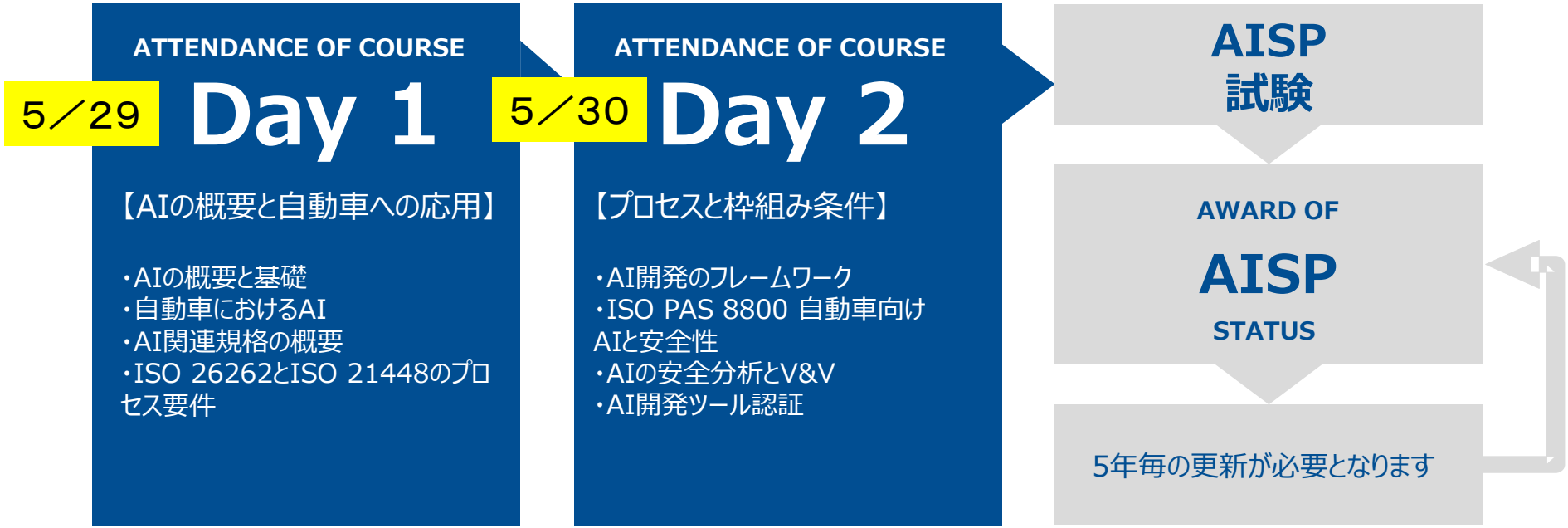
①トレーニング	②簡易 GAP分析	③Q&A ワークショップ	④環境構築 （必要な場合）	⑤プロジェクト 適用	⑥認証
規格トレーニングを受講し、理解を深める。 教育資料	現状プロセスを分析し、課題を抽出する。 GAP分析 レポート カイゼン計画	Q&Aにより、不適合箇所の対策を検討する。 Q&A資料	開発環境を構築し、プロセスを自動化する。 開発・管理システム ツール操作 マニュアル	開発プロジェクトで運用し、結果をフィードバックする。 成果物 プロセス定義書 （更新版）	SGS-TÜV認証、iNTACSアセスメントを行う。 テクニカル レポート 認証書

SGS-TÜV認定AISP資格取得トレーニング

！！開講記念の半額キャンペーン！！

初回の開催については半額で実施いたします。是非この機会をご利用ください。

- AISP：Artificial Intelligence Safety Professional
- 基本規格ISO 26262の補足として、高度自動運転分野における人工知能に関する理解を目的としたトレーニングです。
- 2日間のトレーニングを受講いただくことにより、AISP試験を受験していただくことが可能となります。試験は、最終日の終了後に実施いたします。
  1日目：9:30～17:30、2日目：9:30～16:30（トレーニング）、16:30～17:30（試験）



Thank you!

ご清聴ありがとうございました。

問い合わせ先
Function Safety C&P Connectivity
SGS Japan Inc.

TEL : 050-1780-7876
Eメール: jp.fsafety@sgs.com