ゴール指向要求分析を利用したAIシステムの安全要件の受容性評価
- コミュニティ・バスに搭載した自動運転システムのケーススタディ -

研究コース5　Ai Quality Analysisチーム

研究員： 岩田 正彦 エヌ・ティ・ティ・コミュニケーションズ 株式会社
歌田 真帆 エヌ・ティ・ティ・コミュニケーションズ 株式会社
後藤 優斗 コニカミノルタ 株式会社
高田 晃平 株式会社 東光高岳
柳原 靖司 ブラザー工業 株式会社

主査： 石川 冬樹 国立情報学研究所

副主査： 栗田 太郎 ソニー株式会社
徳本 晋　株式会社富士通研究所

2022/02/25

目次

1. 研究概要

1. 研究概要

AIシステムの社会受容性を考慮して要求分析する枠組み*1の提案

*1 AISA-MVS法
(AI system Safety Analysis method with Multifaceted Viewpoint of Stakeholder)

- システム安全解析技術によりAIシステムの安全性を解析し，解析結果をゴール指向で要求分析することで，受容性の高い要件を選定する
- ケーススタディでは，自動運転システムの認知系AI機能を解析した

カメラ画像からの物体検出，セマンティック・セグメンテーション（Deep Learning）

ケーススタディの対象：自動運転システム

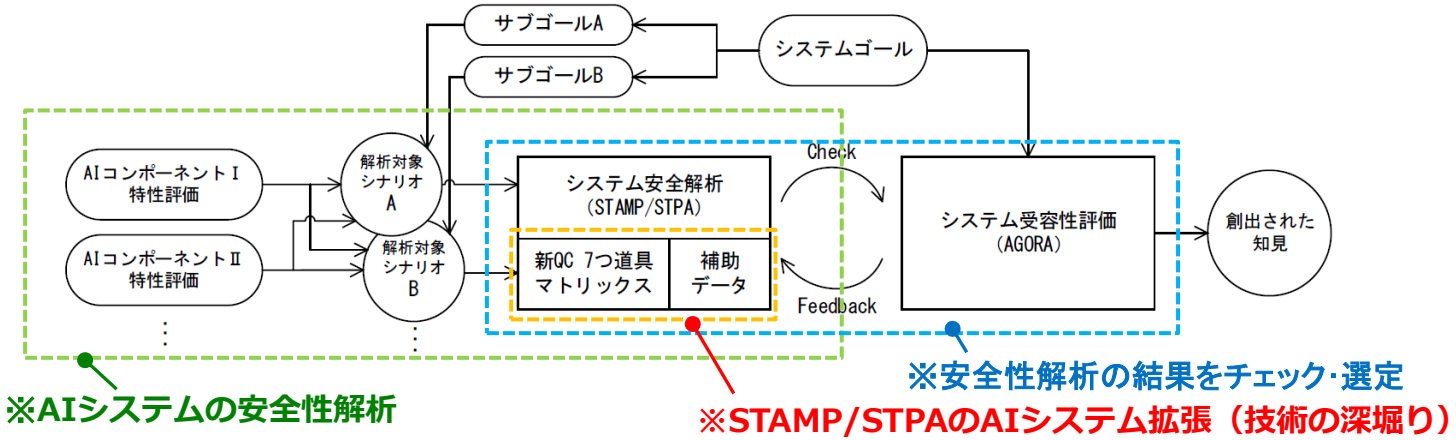


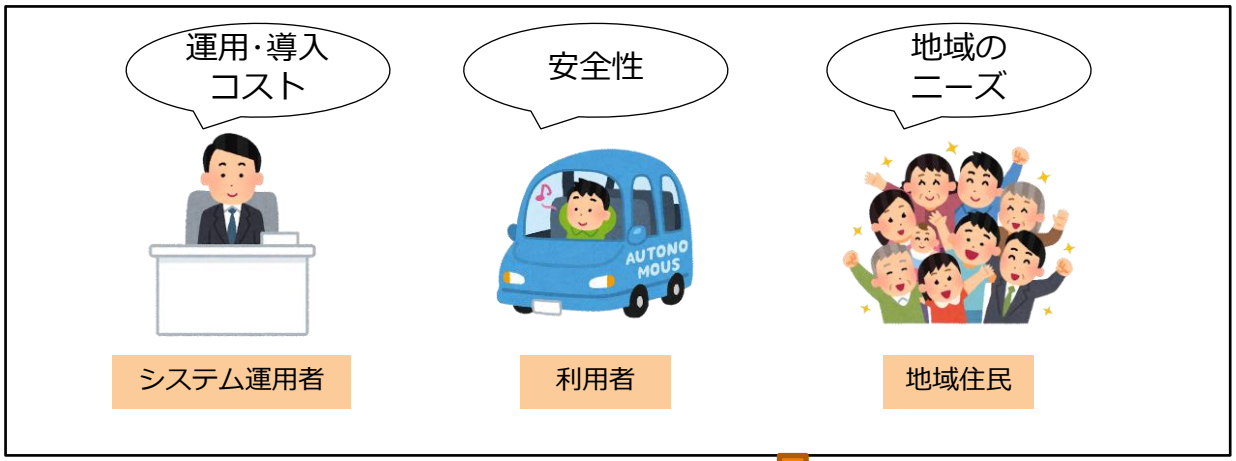
図１ AIシステムの安全解析結果をシステム受容性の観点から評価する枠組み

2. 課題設定

2. 課題設定

■背景と課題

AIシステムの要求分析における課題感：
多様なステークホルダの視点からシステム要求を分析する為の(有識者の知見に依存しない)エンジニアリング手法が確立していない



問題点：多様な視点からの要求(社会受容性)を考慮できる分析手法がない

AIシステムの利用時品質に関して，サービスの社会受容性を考慮しながらステークホルダの納得感を高めるには？

2. 課題設定

■仮説とRQs

目標:
幅広いステークホルダに受け入れられることを考慮した,AIシステムに関わる要件定義の方法を確立する

仮説
AIシステムの安全解析結果に対して，ゴール指向要求分析手法によりステークホルダの志向性を構造的に分析すると，受容性の高い要件を選定できる

RQ1
合意形成が難しいゴールを持つシナリオに対してシステム安全要件の採否判定をAISA-MVS法を使って実施すると，ステークホルダの視点でのシステム要件改善に関する気付きを得られる

RQ2
予めステークホルダの視点を明確にしてシステム安全要件の採否判定を行うと，合意形成が難しいゴールを持つシナリオに対してもステークホルダの受容性を高めやすい

3. 課題解決に向けたアプローチ

3. 課題解決に向けたアプローチ

■提案手法

AIシステムの社会受容性を考慮した要求分析する枠組み

AISA-MVS法

(AI system Safety Analysis method with Multifaceted Viewpoint of Stakeholder)

[特徴]

AI システムの安全要件に対する受容性を評価するためにステークホルダの視点を参考にしながらシステム安全解析の結果を検証し，フィードバックする枠組み

※AGORA満足度行列により，ステークホルダが重視する要件群を特定する

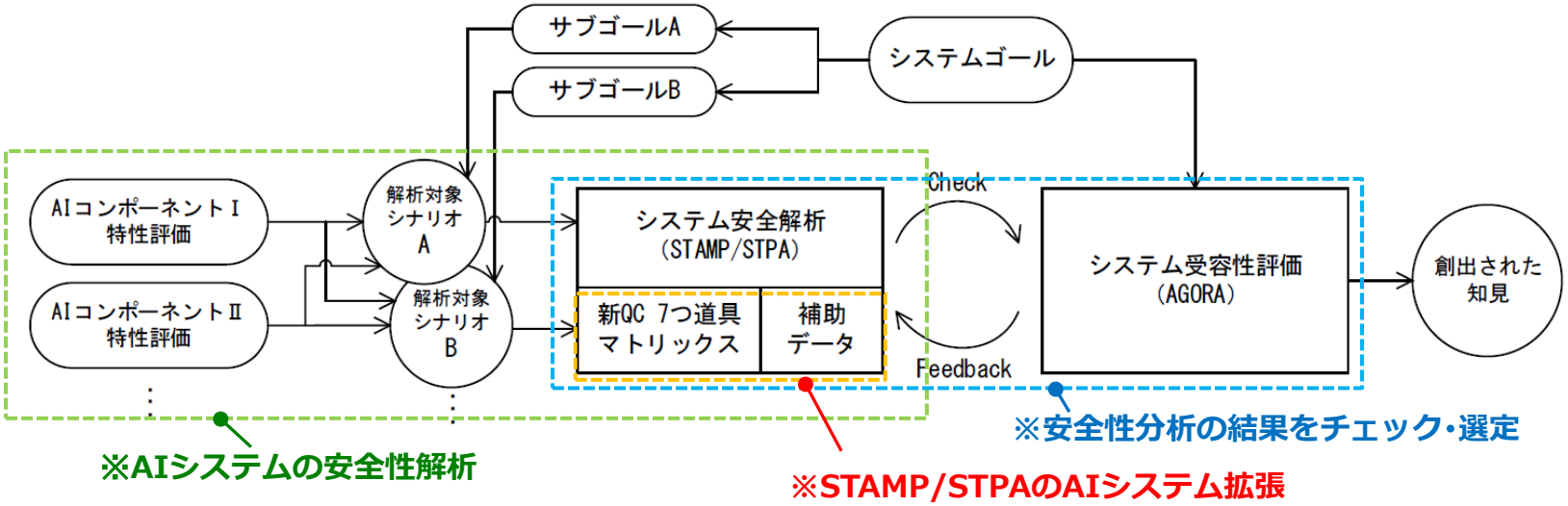


図１ AIシステムの安全解析結果をシステム受容性の観点から評価する枠組み(再掲)

3. 課題解決に向けたアプローチ

■提案手法に関連する技術

AGORA (Attributed Goal-Oriented Requirements Analysis method)

ゴール指向要求分析法の一種で，ツリーグラフを使って主ゴールを分解・詳細化しながら要件を抽出し，満足度行列によりステークホルダの受容性を分析する手法

※本研究ではAIシステムの要件導出と，STAMP/STPA解析結果の受容性検証の目的でAGORAを用いた

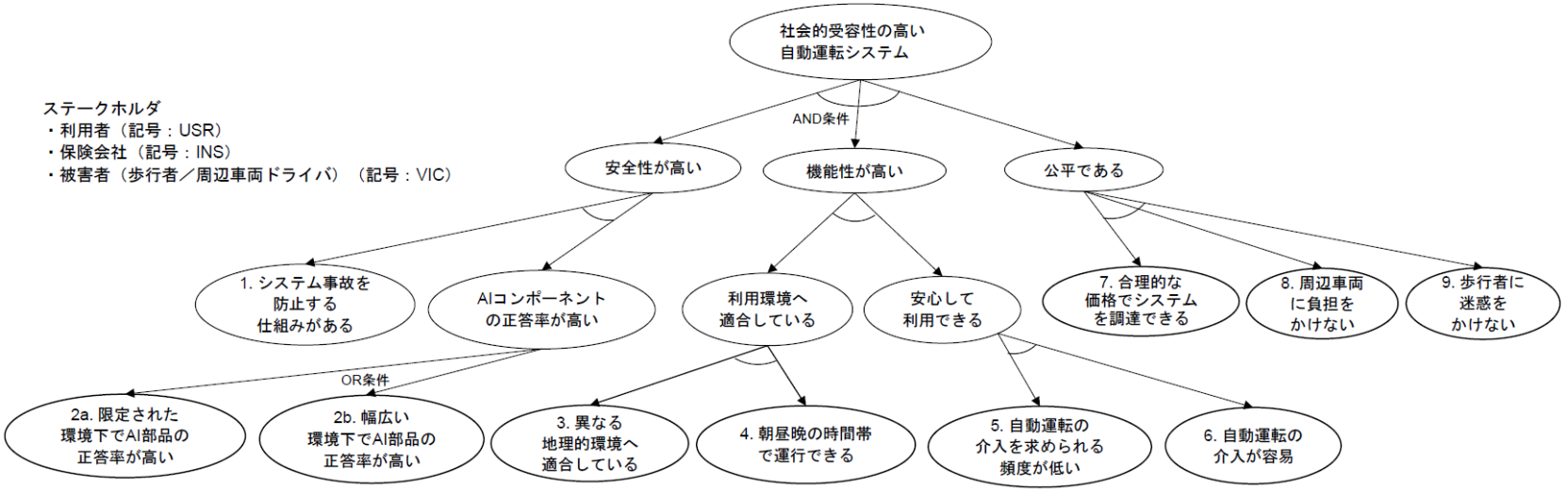


図2 本研究のケーススタディにおける要求分析例

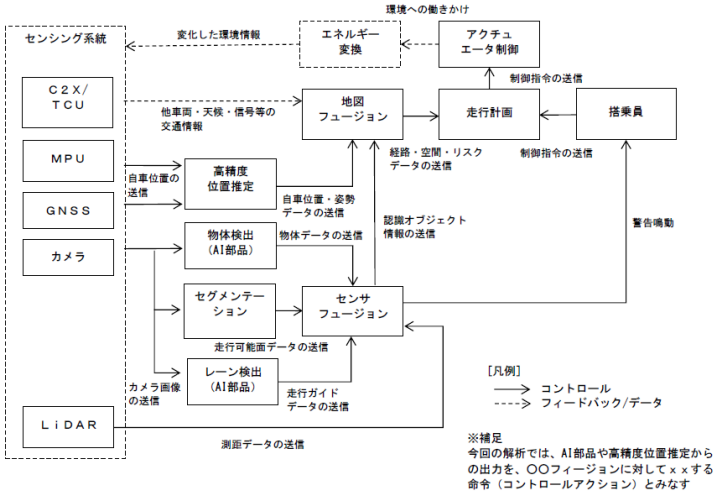
3. 課題解決に向けたアプローチ

■提案手法に関連する技術

STAMP/STPA (Systems-Theoretic Accident Model and Processes/System-Theoretic Process Analysis)

システム設計上のリスクをモデルベースで俯瞰的に解析し，非安全なコントロールアクションに着目して対策を検討する手法
（システム構成要素間の相互作用に着目し，ハザードシナリオを導出する）

※本研究では安全要件を含む AIシステムを解析し，対策案を導出する目的で STAMP/STPAを用いた



UCA	HCF (1)	HCF (2)	HCF (3)
カメラで撮像した画像に映る作業員の解像度が粗い為，物体検出で人として識別されず，アクチュエータ制御（ブレーキ制御）に反映されない（UCA1）	[ヒントワード] コントロールの入力か外部情報が欠けているか間違っている [HCF] カメラ画像の解像度が低い，外光に弱い [対策案] 性能の高いカメラを搭載する	—	—
一時停止標識の検出／バス停に立っている人の検出に失敗し，アクチュエータ制御（ブレーキ制御）に反映されない（UCA2）	[ヒントワード] 生成の欠陥，プロセスの変更，不正確な修正や適応 [HCF] 物体検出モデル生成時の学習データの被覆性や均一性に不備がある [対策案] 学習データを拡充し，再モデル化する	[ヒントワード] 矛盾，不完全，不正確 [HCF] カメラ画像を使った物体検出に依存した標識検出の方式になっている [対策案] C2X/TCUを併用して，周囲の標識情報を取得する	[ヒントワード] 矛盾，不完全，不正確 [HCF] カメラ画像を使った物体検出に依存した方式になっている [対策案] LiDARを併用して，人を検出する

図3 本研究のケーススタディにおけるシステム安全解析例

3. 課題解決に向けたアプローチ

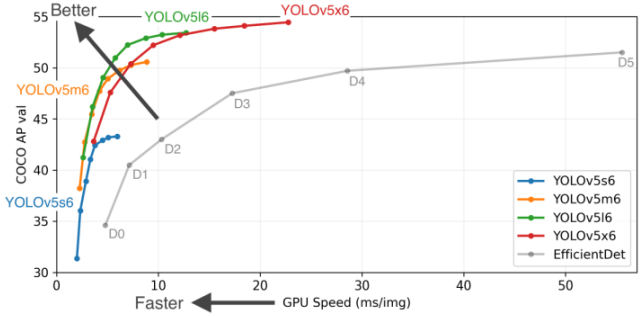
■提案手法に関連する技術

自動運転システムの認知系AI機能

研究用に公開されているアルゴリズムと大規模データセットを利用して
AI部品の特性を調査した

■ 物体検出

アルゴリズム：YOLO v5x
データセット：COCO



■ セマンティック・セグメンテーション

アルゴリズム：
Hierarchical Multi-Scale Attention for Semantic Segmentation
データセット：Cityscapes + Mapillary Vistas

■ 評価データ

バークレー大学が公開しているデータセット
BDD100Kの中から“10K Images”を利用した
https://bair.berkeley.edu/blog/2018/05/30/bdd/

3. 課題解決に向けたアプローチ

■ケーススタディ

自動運転システムのシナリオ

AISA-MVS法によりAIシステムの安全要件の受容性を分析する上で
自動運転レベル３（条件付自動運転）を搭載した乗合いバスを選定

[シナリオA：田舎の乗合いバス]

自動運転システムの利用目的（ゴール）	過疎で高齢化が進む田舎町における移動手段であり，自家用車を運転しないお年寄り等が病院通いや買い物を目的として利用される
走行する道路の特徴	閑散，カーブ，郊外道路
運行時間帯	昼のみ
主要な要求	・移動手段として，地方の小規模自治体が投資可能であることが要求される

[シナリオB：東京2020オリンピックのe-Palette]

自動運転システムの利用目的（ゴール）	大規模祭典における移動手段であり，障碍者を含む多様な方が乗客として想定される
走行する道路の特徴	交差点，中心市街地，施設，施設隣接道路
運行時間帯	運行時間帯は任意（夜間を除く）
主要な要求	・大手スポンサーが運行を支援している為，コスト面の課題はない ・乗客の多様性が高く，道路の構成は複雑で，走行時間帯も幅広い ・複雑な環境下で利用される為，安全を確保する上で難易度が高い

3. 課題解決に向けたアプローチ

■ケーススタディ

自動運転システムの構成と着目ポイント

標準的な自動運転システムの構成例を参考に，STAMP/STPA CS図を作成し，認知系AI機能（赤枠）を取り巻く安全性を解析した

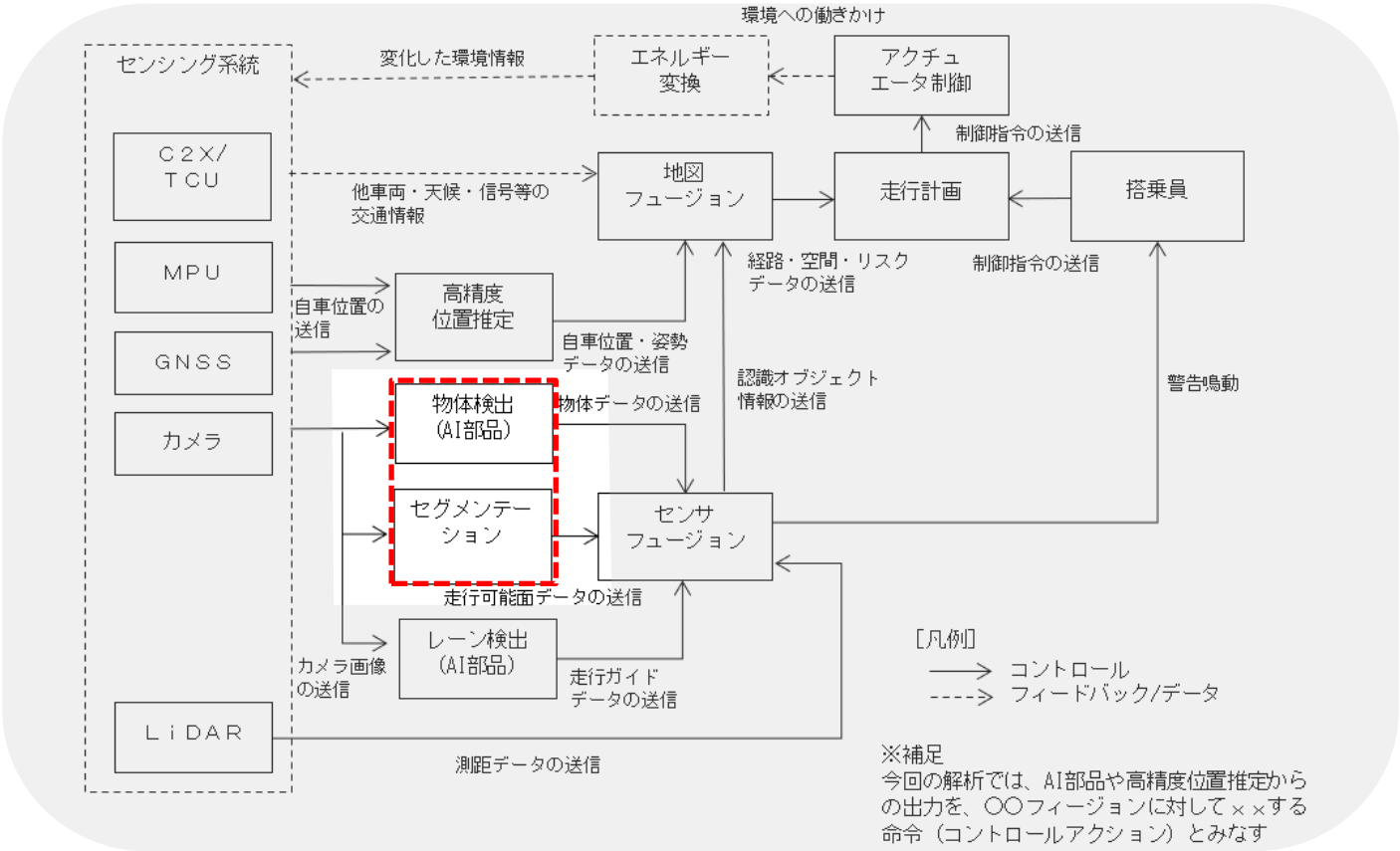
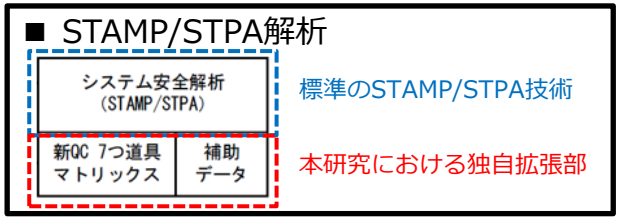


図4 本研究のケーススタディで使用したSTAMP/STPA CS図（システム構成図）

3. 課題解決に向けたアプローチ

■ケーススタディ
STAMP/STPA解析例

環境条件により性能が変動するAIシステムの特徴を踏まえ，QCマトリクスと補助データの態様でAI部品の特性を整理してからSTAMP/STPA解析する



QCマトリックス

	計画（ルート最適化）	予測（リスク軽減・回避）	
AI 部品	ハザード因子		ヒントワード
物体検出 (カメラ部)	走行シーンでは，交差点・中心市街地・駐車場・郊外道路において，ヒヤリハットが 6%前後，中心市街地，郊外道路については，事故に繋がる可能性が夫々0.8%，1.7%程度ある．	事故に繋がる可能性(ASIL-B*2)があるパターンとして，交差点での一時停止標識の見逃し，道路上に立つ作業員の見逃し，目前を走行するトラックの検出精度の低下等がある．	不正確 生成の欠陥 不正確な修正 不正確な適応

*2 ISO 26262 規格で定義されているリスク分類(Automotive Safety Integrity Level)；Vision-based ADAS（ビジョンに基づく先進運転支援システム）における Incorrect Sensor Feedback は ASIL-B に該当．

補助データ

物体検出コンポーネントの時間帯別エラーの出現頻度

[時間帯]	正常	QM	エラーパターンから推定される要求されるASIL			
			A	B	C	D
夜明け,夕暮れ	0.918	0.082	0.000	0.000	0.000	0.000
昼	0.844	0.116	0.036	0.004	0.000	0.000
夜	0.395	0.342	0.263	0.000	0.000	0.000
不明（トンネル，屋内施設等）	0.200	0.600	0.200	0.000	0.000	0.000

※表中の ASIL 値が大きいほど，セルの色を濃くしてある．

ASIL(Automotive Safety Integrity Level):
QM 機能安全を適用しなくてよいレベル，A＜B＜C＜D [厳しい]

自動運転システムのSTAMP/STPA解析例

UCA	HCF (1)	HCF (2)
一時停止標識の検出／バス停に立っている人の検出に失敗し，アクチュエータ制御（ブレーキ制御）に反映されない（UCA2）	ヒントワード：生成の欠陥，プロセスの変更，不正確な修正や適応 HCF：物体検出モデル生成時の学習データの被覆性や均一性に不備がある 対策案：学習データを拡充し，再モデル化する	ヒントワード：矛盾，不完全，不正確 HCF：カメラ画像を使った物体検出に依存した標識検出の方式になっている 対策案：C2X/TCU*3 を併用して，周囲の標識情報を取得する

*3 C2X (Car to X: 車-車間通信や路車間通信のこと), TCU (Telematics Communication Unit:モバイルネットワークを使って自動車の外部との双方向通信を行う装置)

3. 課題解決に向けたアプローチ

■ケーススタディ

STAMP/STPA解析から導出されたハザードシナリオ対策案に対するゴール指向要求分析によるステークホルダ視点での選別

【満足度行列の例】

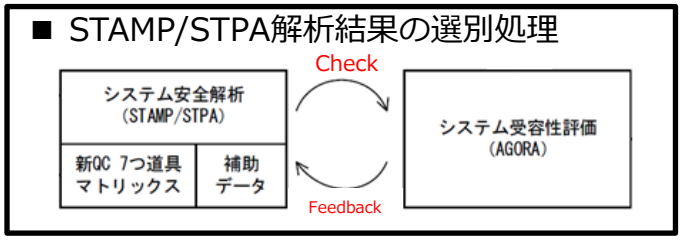
3. 異なる地理的環境へ適合している ★

	USR	INS	VIC
USR	10	9	5
INS	10	9	5
VIC	8	7	6
SUM: 69		VAR: 4.00	

※山間部から市街まで幅広い環境で運行する為，高スコア．

【凡例】　※ステークホルダ
USR 自動運転バスの運行組織
INS 保険会社
VIC 被害者（歩行者等）

■ STAMP/STPA解析結果の選別処理



【ハザードシナリオの選別】

AGORAの満足度行列からステークホルダが重視する要件を特定

ゴール指向要求分析から導出した自動運転システムの要件群

[凡例] ★: 重要要件 1: 有効 0: 関連が小さい -1: 代案で置き換えが必要

UCA	HCF	STAMP対策の代案	1. システム事故を防止する仕組みがある	2a. 限定された環境下でAI部品の正答率が高い	2b. 幅広い環境下でAI部品の正答率が高い	3. 異なる地理的環境へ適合している	4. 朝昼晩の時間帯で運行できる	5. 自動運転の介入を求められる頻度が低い	6. 自動運転の介入が容易	7. 合理的な価格でシステムを調達できる	8.
			★		★	★			★	★	
カメラで撮像した画像に映る作業員の解像度が粗い為，物体検出で人として識別されず，アクチュエータ制御（ブレーキ制御）に反映されない（UCA1）	[ヒントワード] コントロールの入力か外部情報が欠けているか間違っている [HCF] カメラ画像の解像度が低い，外光に弱い [対策案] 性能の高いカメラを搭載する	—	0	N/A	1	1	N/A	N/A	0	0	
一時停止標識の検出／バス停に立っている人の検出に失敗し，アクチュエータ制御（ブレーキ制御）に反映されない（UCA2）	[ヒントワード] 生成の欠陥，プロセスの変更，不正確な修正や適応 [HCF] 物体検出モデル生成時の学習データの被覆性や均一性に不備がある [対策案] 学習データを拡充し，再モデル化する	—	0	N/A	1	1	N/A	N/A	0	0	
	[ヒントワード] 矛盾，不完全，不正確 [HCF] カメラ画像を使った物体検出に依存した標識検出の方式になっている [対策案] C2X/TCUを併用して，周囲の標識情報を取得する	（実施しない） ※田舎なので必要性が乏しい	1	N/A	0	0	N/A	N/A	0	-1	

STAMP/STPA ハザード対策案

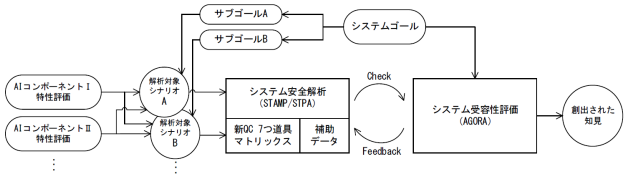
縦横の突合せにより，ステークホルダの受容性の観点からハザード対策案の採否を決める

4. 評価（実験）

4. 評価（実験） 実験概要

■予備実験

- **基準データの作成**
  AISA-MVS法により，自動運転バスのシステム安全要件を作成する
  ※ステークホルダの受容性観点よりシナリオA,Bに対し各々行う



[基準データ]
ステークホルダの受容性を考慮したシステム安全要件

- **予備実験**
  シナリオA,BのSTAMP/STPAの解析結果を被験者に見せ，被験者の経験･勘でシステム安全要件の採否をしてもらう

■本実験

- **本実験1**
  被験者へ「基準データ」を見せ，最初に各自が採否判定した結果を訂正するか否かを考えてもらう
- **本実験2**
  被験者へ「基準データ」を見せ，ケーススタディで想定したステータスホルダの受容性をどの程度満足できているかを聞き取る

■予備実験の結果

自動運転システム搭載バスのシステム安全要件について，シナリオA（田舎の乗合いバス）の方が，シナリオB（大規模祭典のコミュータ）よりも，システム受容性に対する被験者の捉え方のばらつきが大きいことが分かった

⇒ ステークホルダの合意形成が難しい

項目	シナリオA	シナリオB
基準データにおける却下数（AISA-MVS 法）	7 件	1 件
被験者の平均却下数	6.4 件	5.4 件
基準データに対する被験者判定結果のばらつき（分散）	25.1	12.2

図5 ハザードシナリオ対策案の被験者回答結果の概要

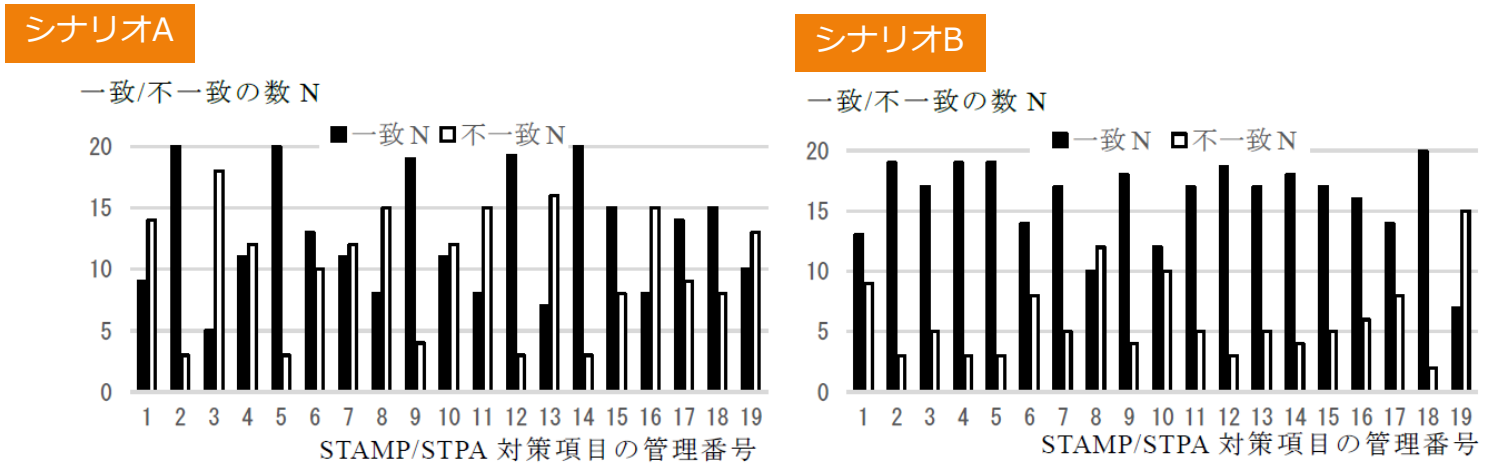


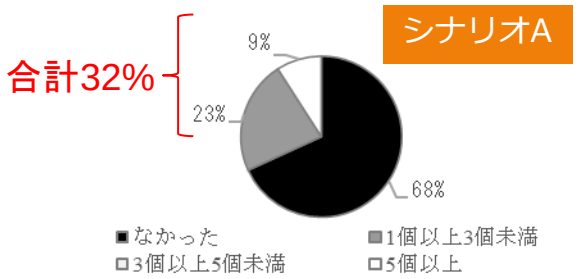
図6 ハザードシナリオ対策案の基準データと被験者回答結果の一致･不一致の数

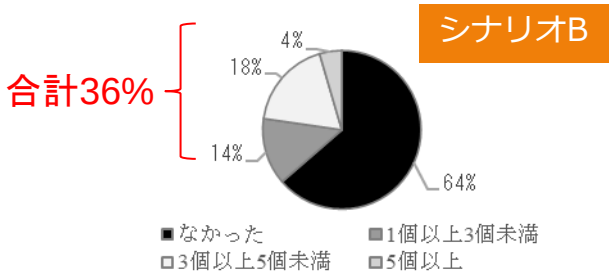
4. 評価（実験） 実験1,実験2の結果

■実験1の結果

被験者へAISA-MVS法由来の「基準データ」を見せた後，経験･勘でハザードシナリオの採否判定してもらった結果，3割の被験者は１個以上訂正したことが分かった

⇒ステークホルダの視点を入れるとシステム要件改善に関する気付きを得られる

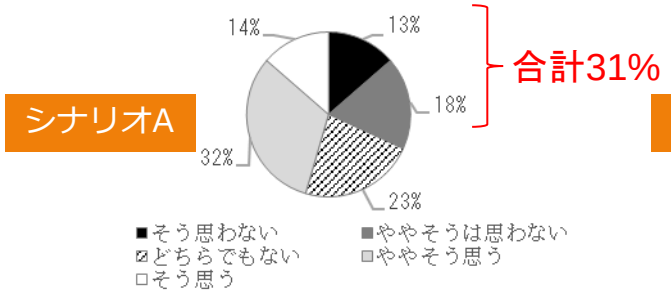


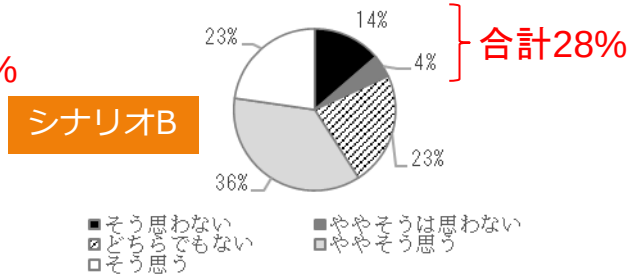


■実験2の結果

被験者へAISA-MVS法由来の「基準データ」がステークホルダの受容性を満足しているか調査した結果，否定的な意見は，3割程度であることが分かった

⇒提案手法でシステム安全要件を導出すると，ステークホルダの受容性が高くなる





5. 考察

5. 考察

■得られた知見

- AIシステムの安全解析より導出した要件の採否判定を行う際，
  AISA-MVS法を用いるとステークホルダの視点を考慮できる
  ⇒ 運用フェーズにおけるシステムの受容性を高めることに寄与できそう
  （※ケーススタディには前提条件がある為，一定の効果が確認された状態）

  【主な前提条件】
  - ◆シナリオで用いた乗合いバスの環境条件は単純化されている
  - ◆ステークホルダに対して，研究員の想定が含まれている

■ AISA-MVS法の実用性

- 現場活用での再現性という点では，ベースとなっている技術が，
  既知の実績があるものであり，取り扱いが容易である
  ⇒ 誰でも使える

6. 成果

6. 成果

■研究課題に対する成果

- AIシステムの安全解析結果をゴール指向要求分析により検証する枠組みとしてAISA-MVS法を考案し，ステークホルダの受容性を考慮したハザードシナリオ対策案の選別を可能にした

■副次的な成果

- (自動運転システムに搭載されるような)環境条件によって性能が変動するAIコンポーネントをSTAMP/STPAにより解析する際，QCマトリックスや補助データを活用することで解析精度を高めた

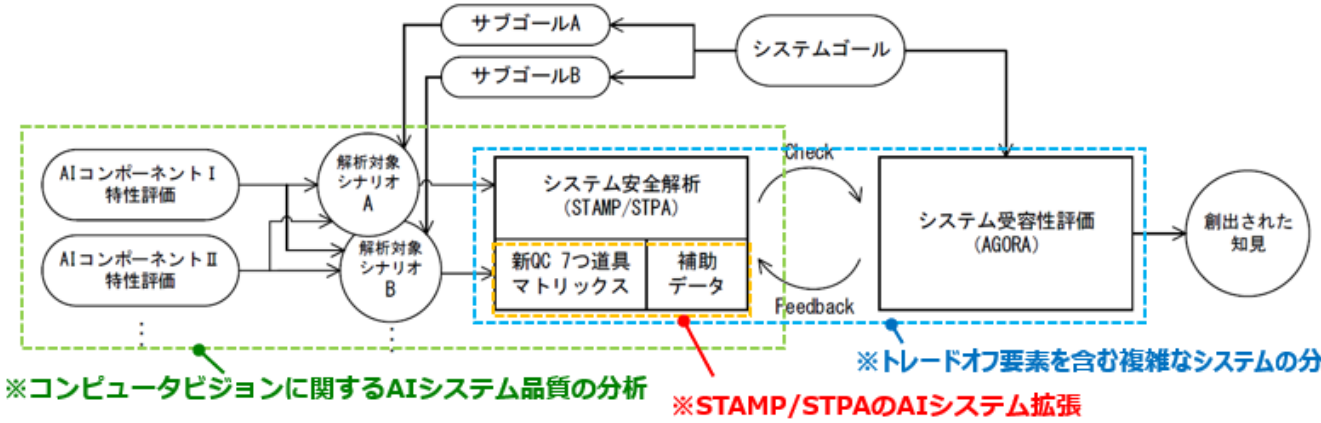


図１ AIシステムの安全解析結果をシステム受容性の観点から評価する枠組み（再掲）

ご清聴ありがとうございました