

AIとシステムセーフティとの関係 －規格適合性において何が課題か－

森 崇
鉄道認証室

講演内容

1. 装置による安全と人の役割分担
2. 機能安全とは
3. 機能安全とは異なる開発サイクル
4. 機能安全規格上の安全に関する設計のポイント
5. AI活用の方針案
6. もう一つの規格上の課題

経歴

1992年	西日本旅客鉄道株式会社入社	岡山信号通信区配属		
1993年	株式会社京三製作所出向			
1994年	西日本旅客鉄道株式会社	大阪信号工事所・京都信号工事所で工事担当		
1997年	西日本旅客鉄道株式会社	技術開発推進部(現イノベーション本部)		
	2024年7月まで配属(2021年より休職)			
	無線式列車制御、IP network, ソフトウェア定義無線など担当			
2019年	独立行政法人自動車技術総合機構	交通安全環境研究所	鉄道認証室	客員専門調査員委嘱
2021年	独立行政法人自動車技術総合機構	交通安全環境研究所	鉄道認証室	任期付職員(主席研究員)
2024年	西日本旅客鉄道株式会社退職、 独立行政法人自動車技術総合機構	交通安全環境研究所	鉄道認証室	主席鉄道認証研究員

鉄道関係の国際規格のうち、製品認証業務を行っている。

0. 登場人物

鉄道信号メーカー カバ興業チーム



カバ興業 社長
座右の銘：技術と直感



カバ興業 営業 カバお
「怒られてナンボの毎日」



カバ興業 技術 オタかば
「面白くなければ技術じゃない」



カバ興業 プログラマ
ハッキングカバ
「俺しかできないことをやる」



カバ興業設計課長 カバ実
「全体のレベルアップ」

鉄道事業者 カバ鉄道チーム



カバ鉄道 社長
品格の経営、根拠ある経営



カバ鉄道 電気課長
口癖：安くてエエもん持って来い！



カバ鉄道 乗務員 カバどん
いつかは自動化されるかも。ドキドキ



AIカバどんくん
カバどんの夢に出てきた変なやつ



謎な奴

謎のフリーコンサル
なぞカバ
「知識は力！」

1. 装置による安全と人の役割分担

カバで頑張る！



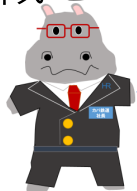
カバ鉄道社長

安全のかなめは乗務員！カバどん君！君の双肩にかかっている！
経験を積んで、しっかり頼むぞ！経験が安全につながる。いろいろ試してみて、ベストを生み出してくれ。お願いしますぞ。



分かりました。社長。精一杯やります。。

機械との協調



安全のかなめは乗務員！カバどん君！君の双肩にかかっている！しかし、だな、カバはミス
をすることだってあるだろう。そのために保安装置がある。少なくとも前の列車に衝突するこ
とは機械で防いでいる。



しかし快適な運転、その時々状況に応じた運転は君しかできない。そこをお願いしたい！

分かりました。社長。精一杯やります。。

2. 機能安全とは

IEC 61508-4 Functional safety(機能安全)

part of the overall safety relating to the EUC and the EUC control system that depends on the correct functioning of the E/E/PE safety-related systems and other risk reduction measures

EUC及びEUC制御系の全体に関する安全のうち、E/E/PE安全関連系及び他リスク軽減措置の正常な機能に依存する部分。



こんな難しいことゆうてどうなるんや。。わかるように説明するのがコンサルやろ。
あそうか、わざと難しい言い方して、説明するのにまたカネをとる。エエ商売やのう。そんな商売してまっとうな
カバ興業からカネとるんか。キタナイのう。。



- 「正常な機能により、対象のシステムのリスクを軽減する」のが、機能安全と言えます。



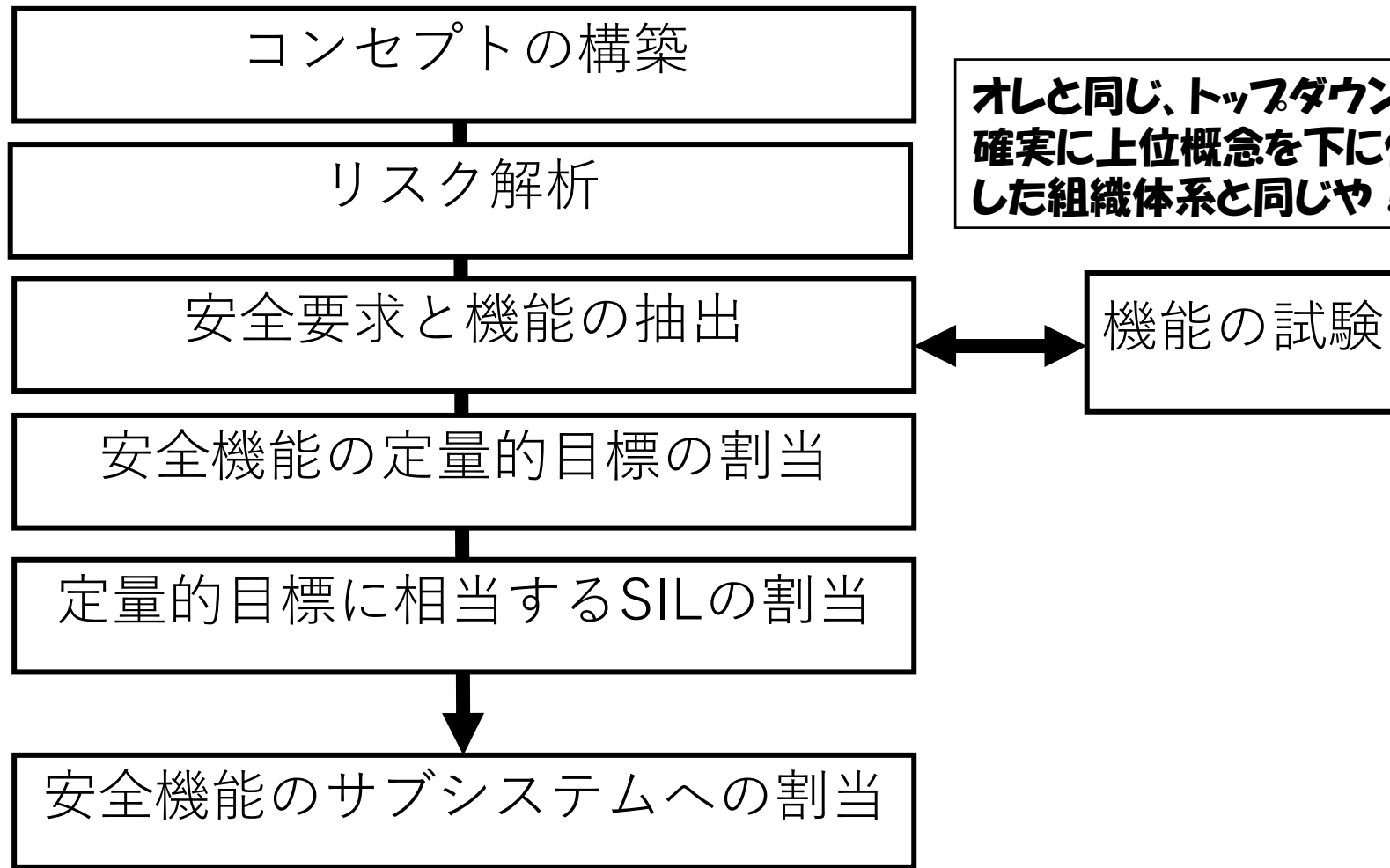
正常な機能とはなんや！壊れたら正常じゃないやろ！正常じゃないものに正常を要求するってどうすんねん。



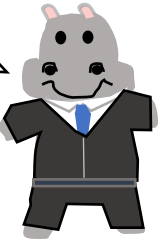
故障を事前に解析して、故障しても安全側に遷移するような手法で、安全を保つ方法ともいえると思います。
すなわち、設計の段階から、低減すべき故障の状態を把握し、その把握された故障は、設計通りの状態に遷移させるといえると思います。



2. 機能安全とは



オレと同じ、トップダウンアプローチだな。
確実に上位概念を下に伝えていく。まさにしっかりした組織体系と同じじゃ！



2. 機能安全とは 重要な実現方法[fail-safe]

Area Signalling and security apparatus for railways / General concepts and terms

EV ref 821-01-10

en **fail-safe**, adj
able to enter or remain in a safe state in the event of a failure

fr **de sécurité intrinsèque**, adj
capable d'entrer ou de rester dans un état sûr dans l'éventualité d'une défaillance

ar عطل آمن

de sicher bei Ausfall, adjektivisch
"fail-safe", adjektivisch

es seguridad intrínseca

it sicurezza intrinseca

ko 고장 안전, 형용사

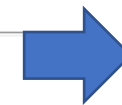
ja フェールセーフ

pl uszkodzony bezpieczny, adj
fail-safe, adj



故障において安全状態
に遷移または居続ける
こと

IEC 60050に
フェールセーフの定
義もあるんや

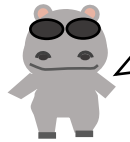
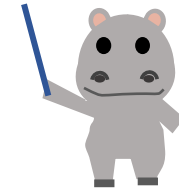


そのまんま？もう少し
説明欲しいね。

2. 機能安全とは 重要な実現方法 detection and negation



リスクアセスメントにより、
機能を定義し作りこみ

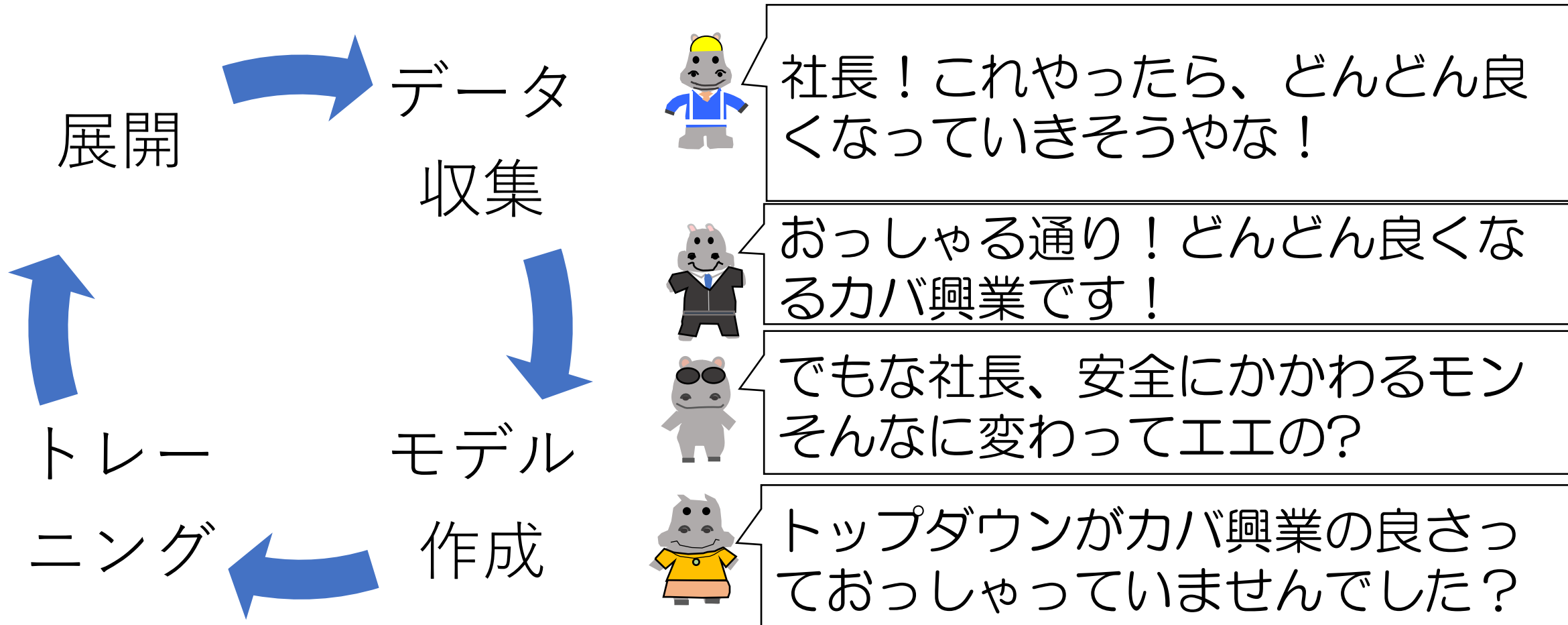


壊れたことが検知できなかったら。。ガクブルやな。



検知できても、止められなかったら意味ないよね。

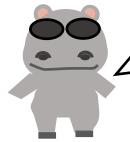
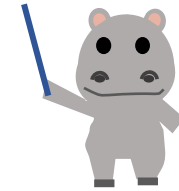
3. 機能安全とは異なるアジャイルの開発サイクル



3. 前提が変化する場合のdetection and negation



学習によってできるできないが異なる。。。。



壊れたことが検知できなかったら。。ガクブルやな。



検知できても、止められなかったら意味ないよね。

4. 機能安全とAI活用の開発サイクルの差異

機能安全規格の考え方

- リスク解析で抽出され、低減すべきと決めた事項は、低減安全機能を定義し、その機能を設計に展開する。
【トレーサビリティ】
- 安全機能は、明確に動作定義が要求されており、決定論的な状態遷移が前提となっている。

AIなどの考え方

- 開発の途上において、PoC(Proof of Concept)を行い、プロトタイプを実証することにより開発を進めていく。→これはプロトタイプがそのまま実現されることもあるが、規格上は制約が多い。
- 安全機能について、設計時に想定されていない事柄も対策できることもあるが、学習データに大いに依存する。

適材適所ってあるしな。。



4. 機能安全とAI活用の開発サイクルの差異

機能安全規格の考え方

- 学習によって状態遷移確率が変わり、学習により状態遷移が決定するような考え方には立っていない。

鉄道の安全関連系ソフトウェア規格IEC 62279の規格上は、学習によってのデータ生成は“Application data and algorithms”として位置づけられ、そのデータの試験と確認し、リリースバージョンを確定する。変更の場合は所定の試験と確認が再度必要。

AIなどの考え方

- 当然ながら、学習ベースによるアプリケーションアルゴリズムの設定が前提である。
- 決定論的な状態遷移という考え方より、確率論的な状態遷移。
- 実運用後もブラッシュアップも考慮されている。この場合は運用中に状態遷移確率が変わりうる。

予想通りかなり考え方が異なるな。



4. 機能安全規格上の安全に関する設計のポイント

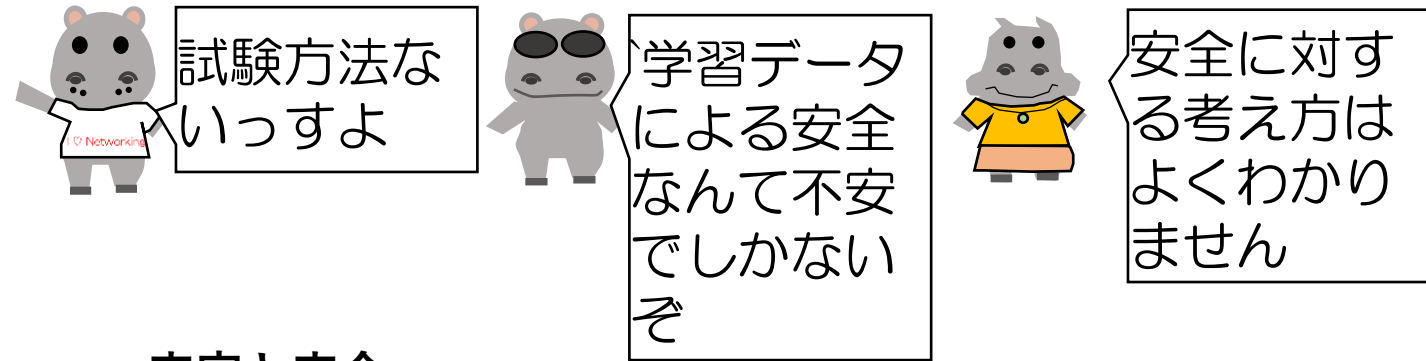
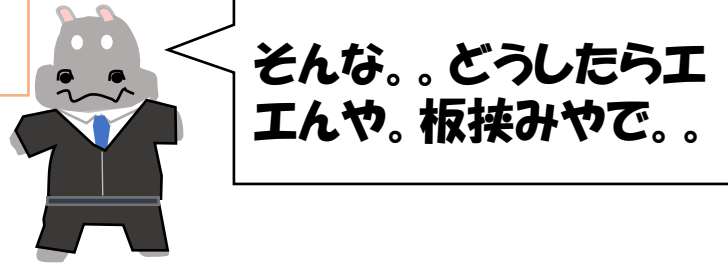
- IEC 61508には、Artificial intelligence fault correctionの是非について述べられており、安全性インテグリティ(SIL)1以外の安全性を要求される場合は、Not Recommended (NR)となっている。
- IEC 62279 においても、Artificial intelligence fault correctionの是非について述べられており、ソフトウェア安全性インテグリティ(SSIL)0以外の安全性を要求される場合は、Not Recommended (NR)となっている。

5. AI活用の方針案



Team勢い第一

Team板挟み



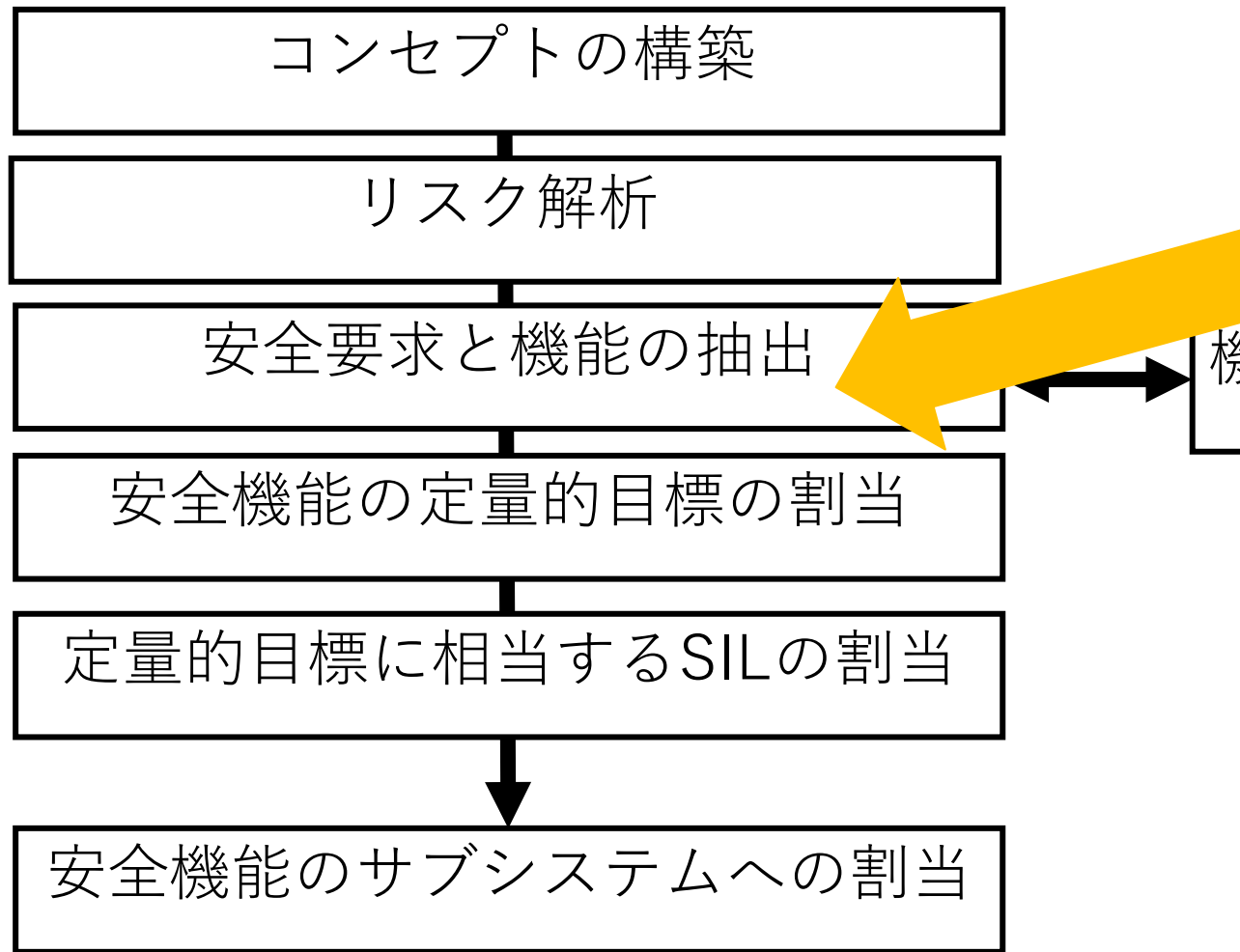
Team安定と安全

5. AI活用の方針案

NISTの AI Risk Management Framework (AI RMF)及びその解説であるRMF Playbookをもとに調査

リスク概要	具体的な鉄道への影響
リスクにおいては重篤度×頻度であるため、評価と対処基準が必要 (GOVERN 1.3)	Random Fault以外における確率論アプローチの考え方の整理
運用後の経時的変化による安全度の変化 (MANAGE 2.2)	運用中に学習することにより安全度が考えていたものと異なる可能性
第三者データ、ソフトウェアなどによる安全性の担保 (MANAGE 3.1)	Validation方法の確立
用意したデータの品質の問題による安全性の影響 (MANAGE 3.2)	学習させるデータの品質の見極めと安全性が直結する
学習の複雑さの増加と再現性の問題 (MANAGE 4.2,4.3)	不具合が起こった場合、再現試験が求められるが、再現されるとは限らない
決定ポイントの複雑性による説明の難しさおよび試験方法の標準化不足 (MANAGE 2.2)	規格上試験のカバレッジと要求とのトレースが求められるだけでなく、経時的な機能変化について対処が必要

5. AI活用の方針案



対策としての安全機能をできるだけシンプルなものとする。例えば、全線区にわたり適正な速度で自動運転を行う。という機能に安全性を持たせるのではなく、危険な速度に到達した場合、停止させるなどシンプルな安全機能とし、制御全体に安全機能を持たせない。安全機能以外は規格上AIを活用しても問題ない。

5. AI活用の方針案

AIで頑張る！



安全のかなめは乗務員！AIカバどん君！君の双肩にかかっている！
経験を積んで、しっかり頼むぞ！経験が安全につながる。いろいろ試してみて、ベストを生み出してくれ。お願いしますぞ。



分かりました。社長。精一杯やります。。

機能安全装置との協調



安全のかなめは乗務員！AIカバどん君！君の双肩にかかっている！しかし、だな、カバはミスをする事だってあるだろう。そのために機能安全による保安装置がある。少なくとも列車に衝突することはほかの装置で防いでいる。

しかし快適な運転、その時々状況に応じた運転は君しかできない。そこをお願いしたい！



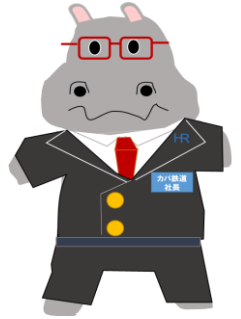
ドンドン学習していきますよ！

5. AI活用の方針案



社長。。私はクビですか。。

はあ、AIカバどん君？それはなんや。変な夢見たんちゃうか？



- 安全機能の割り当てをできるだけ少なくする。
 - システム全体に割り当てるのではなく、機能がFailした場合のexternal countermeasuresとして位置づける。これは機能安全規格IEC 61508の大原則としてもマッチしている。
 - こうすることにより、きめ細かな制御はAIに割り当て、シンプルな安全機能はこれまでの決定論的な機能で位置付けることができる。
 - 今人間が行っている業務を置き換え、安全を最後担保する装置は置き換えの対象外とすると、簡単に考えることができる。

6. もう一つの規格上の課題

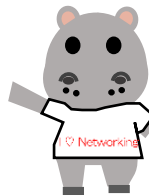
機能安全規格の考え方

ハードウェアの故障と安全性

→ランダム故障として、コンポーネントの故障率を組み合わせ、危険側故障頻度を推定

ソフトウェアなどの故障と安全性

→システマティック故障として、管理手法、試験を厳密にすることにより対処



Tolerable Hazard Rateとの関係性の整理が必要ですね。。。

AIなどの考え方

ハードウェアの故障と安全性

→ランダム故障として、コンポーネントの故障率を組み合わせ、危険側故障頻度を推定

Generic softwareなどの故障と安全性

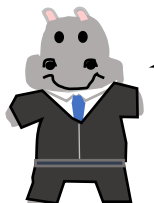
→システマティック故障として、管理手法、試験を厳密にすることにより対処

アプリケーションデータ故障と安全性

→ニューラルネットの状態遷移確率など、確率論的なアプローチが必要

7. まとめ

- 機能安全における安全確保の方法と、アジャイル的な開発方法について、対比を行い、その思想の違いをご説明しました。
- AI活用において、仕様と試験の網羅性、経時的な変化についての対応が必須ですが、規格上の解決策はまだ述べられていません。
- 当面は、できるだけ安全機能を簡単にし、それ以外の機能のAIの活用を進めてはいかがでしょうか。



もっといい考え持っている奴、来たれカバ興業へ。高給優遇するで！



これは社長の嘘をついている目です。騙されないで。。