

装置による安全と人の役割分担はどちらが適正でしょうか？まずAIは何を目指すべきでしょうか？

カバ鉄道社長

安全のかなめは乗務員！カバお君！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。

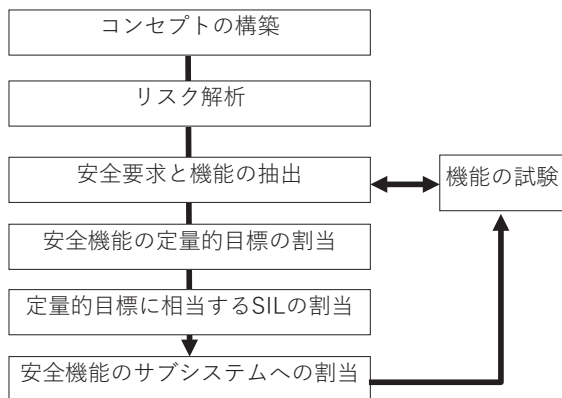
分かりました。社長。精一杯やります。。



安全のかなめは乗務員！カバお君！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。。

一般的な機能安全における設計の進め方



機能安全規格上の安全に関する設計のポイント

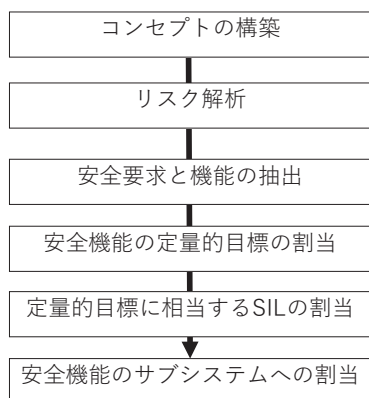
- ・リスク解析により抽出され、低減すべきと決めた事項は、低減するための安全機能を定義し、その機能を設計に確実に展開する。【トレーサビリティの確保】
- ・安全機能は、明確に動作定義が要求されており、決定論的な状態遷移が前提となっている。
- ・学習によって状態遷移確率が変化し、学習により状態遷移が決定するような考え方には立っていない。
- ・IEC 61508には、Artificial intelligence fault correctionの是非について述べられており、安全性インテグリティ(SIL)1以外の安全性を要求される場合は、Not Recommended (NR)となっている。
- ・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。

AI活用における主な課題と影響

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

当面の活用の方策



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。

最悪の事態を避けるのは、誰が何と言おうと、すぐ止めるロジックだ。経験工学ではない。これは決定論的なロジックで行く！しかし、乗り心地や、急速な減速を避け、できるだけスムーズに異常事態に対応できるというのは、学習していく必要があると考える。

システム全体

安全機能；
決定論的論理

非安全機能；
決定論的論理でなくとも可能

社長。。私はクビですか。。

何を言っているのだ。AIにいろいろ教えてやるのは、カバお、お前しかいないし、新たな価値はカバからしか生まれないぞ！

将来的には、運用後のドリフトの考え方が規格上も定まり、一定のTolerable Functional Failure Rateにライフサイクル全体で収まる見込みがはっきりした際には、さらなる活用も可能であると考え。