資料１

人工知能(AI)と安全
ー標準化の観点からー

2019年2月5日
ナブテスコ（株）技術本部 電気電子エンジニアリング部
佐藤吉信

演者紹介

佐藤吉信：現在、IEC/TC 56国内委員会委員長、IEC 61508改訂エキスパート、IEC 63187開発エキスパート、ナブテスコ(株)技術本部電気電子エンジニアリング部技術顧問（工博）

- 東京商船大学/東京海洋大学教授、(株)日本環境認証機構機能安全担当部長を経て、2017年4月より現職。
- 1994年より IEC 61508審議国内委員会幹事/主査及び規格開発国際エキスパート歴任。
- 2004年 システム安全及び機能安全の教育・研究の功績により電子通信学会フェロー。
- 2008年 経済産業省よりディペンダビリティ/機能安全に関する国際標準化貢献者賞受賞。
- 2012年システム安全/機能安全の学術/産業界への貢献により安全工学会北川学術賞受賞。
- システム安全/リスクアセスメント/機能安全に関する論文・著書を多数公刊。

講演内容

- 科学・技術システムが関連する安全の標準に係る基本原理
  - リスクと安全
  - リスクマネジメントプロセスと機能安全プロセス
- 電気・電子・プログラマブル電子安全関連系の機能安全
  - 偶発的/決定論的不具合原因と安全機能の不全
  - 機能安全のフレームワーク
- ソフトウェア(S/W)の安全性
  - 安全なS/Wとは？－S/Wの安全要求事項
  - 人工知能(AI)は安全か？
- AIを実装したシステムの安全
  - AIの安全性確保
  - 多重防護層によるシステムの安全確保（介護ロボット、自動運転車を事例として）
  - 機能安全達成の責任と実施者

リスクとは？ー国際標準による定義

ISO 31000:2009　リスクマネジメントー原則及び指針(JIS Q 31000:2010)

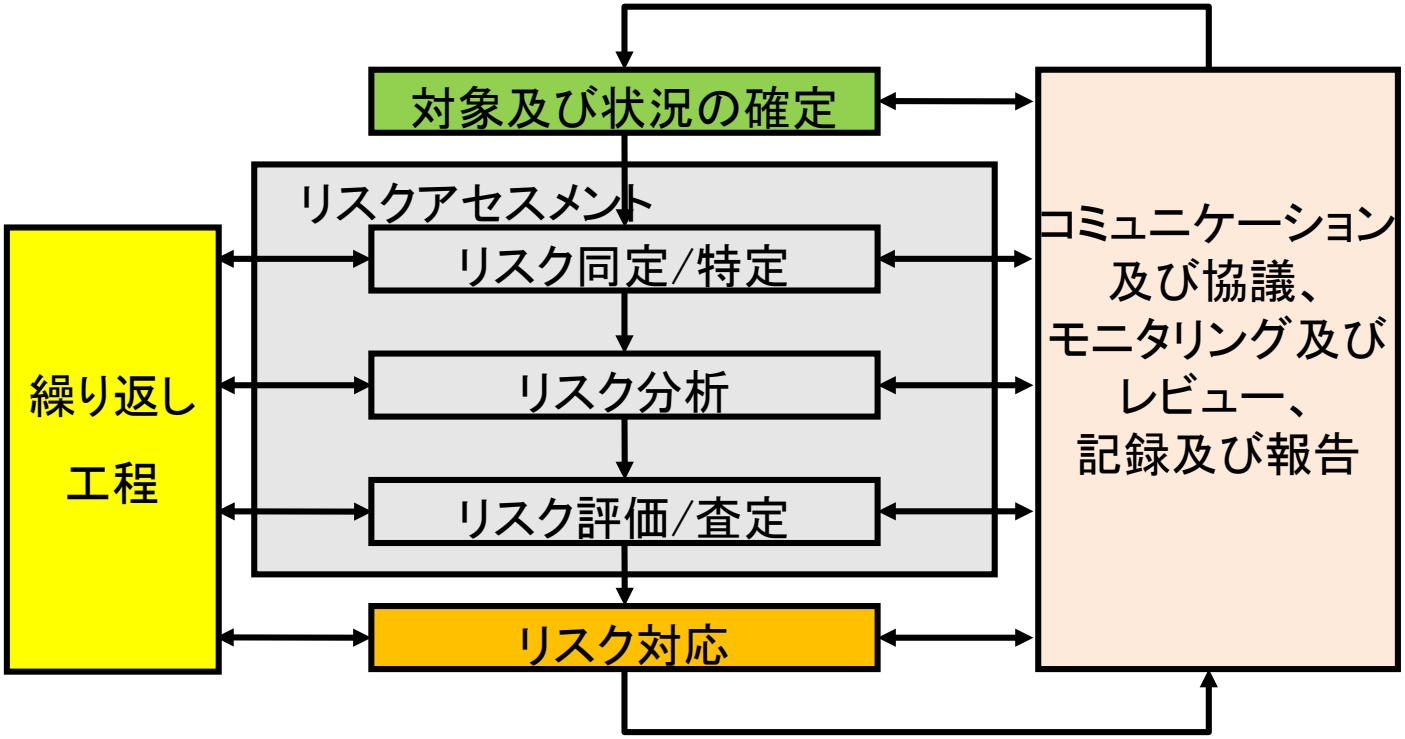
- リスク：目的に対する不確かさの影響（及びその結果）
  - 影響（及びその結果）ー期待されていることからの乖離（逸脱）
  - 乖離（逸脱）には、好ましい(Positive)及び/又は好ましくない(Negative)ものがある。
  - リスクは、しばしば潜在的に起こり得る事象およびその結果（状態）、すなわち事象とその結果の組合せで表現される（特に安全の分野）。

安全関連リスク及び安全とは？ー国際標準による定義

ISO/IEC Guide 51:2014　安全側面－規格への導入指針(JIS Z 8051：2004)

- 安全関連リスク：危害の発生の確率及びその危害の重篤度の組合せ
- 危害：人の受ける身体的傷害若しくは健康傷害、又は財産若しくは環境の受ける害
  - 危害の発生は危険事象、危害は結果（状態）である。
  - 危害事象は好ましくない事象である。
  - 安全の目的は危害がないことであり、安全性に特化したリスクの定義である。
- 安全：許容できない安全関連リスクから免れていること
- 安全を達成する仕組み
  - リスクの除去（リスク源の除去）による→固有（本質）安全
  - 発生確率を低減/重篤度を軽減する安全機能による→機能安全
    - ✓ 電気・電子技術を実装した安全機能を履行するアイテム→電気・電子・プログラマブル電子安全関連系

リスクマネジメントのプロセス



出典:ISO 31000:2009　リスクマネジメントー原則及び指針(JIS Q 31000:2010)

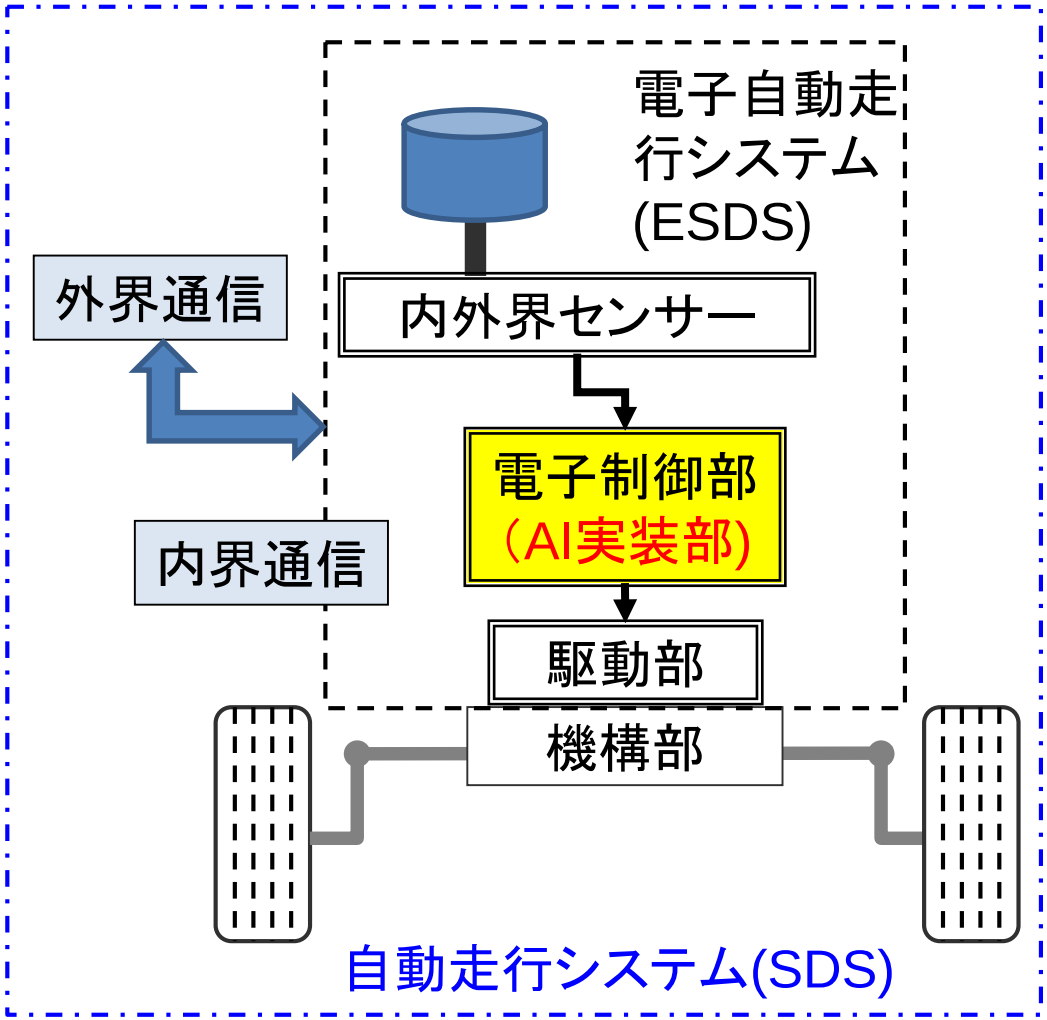
想定内と想定外事象のリスクと機能安全

知識上及び対策の想定内・想定外

事象及び原因を知っていたか（知識上）？ 事象への対策を超える影響か（偶発性）？	知識上の観点（Epistemic）	
	想定内事象 （既知の）	想定外事象 （未知の）
対策の想定内の影響 （対策で対応可） （Aleatory）	制御された リスク	抑制された メタリスク
対策の想定外の影響 （対策の範囲外） （Aleatory）	既知の 残存・残留リスク	危険な メタリスク

出典：佐藤吉信、機能安全の基礎、日本規格協会、27−29、June 2014

電気・電子・プログラマブル電子安全関連系とは？
ー自動運転における電子自動走行システムー



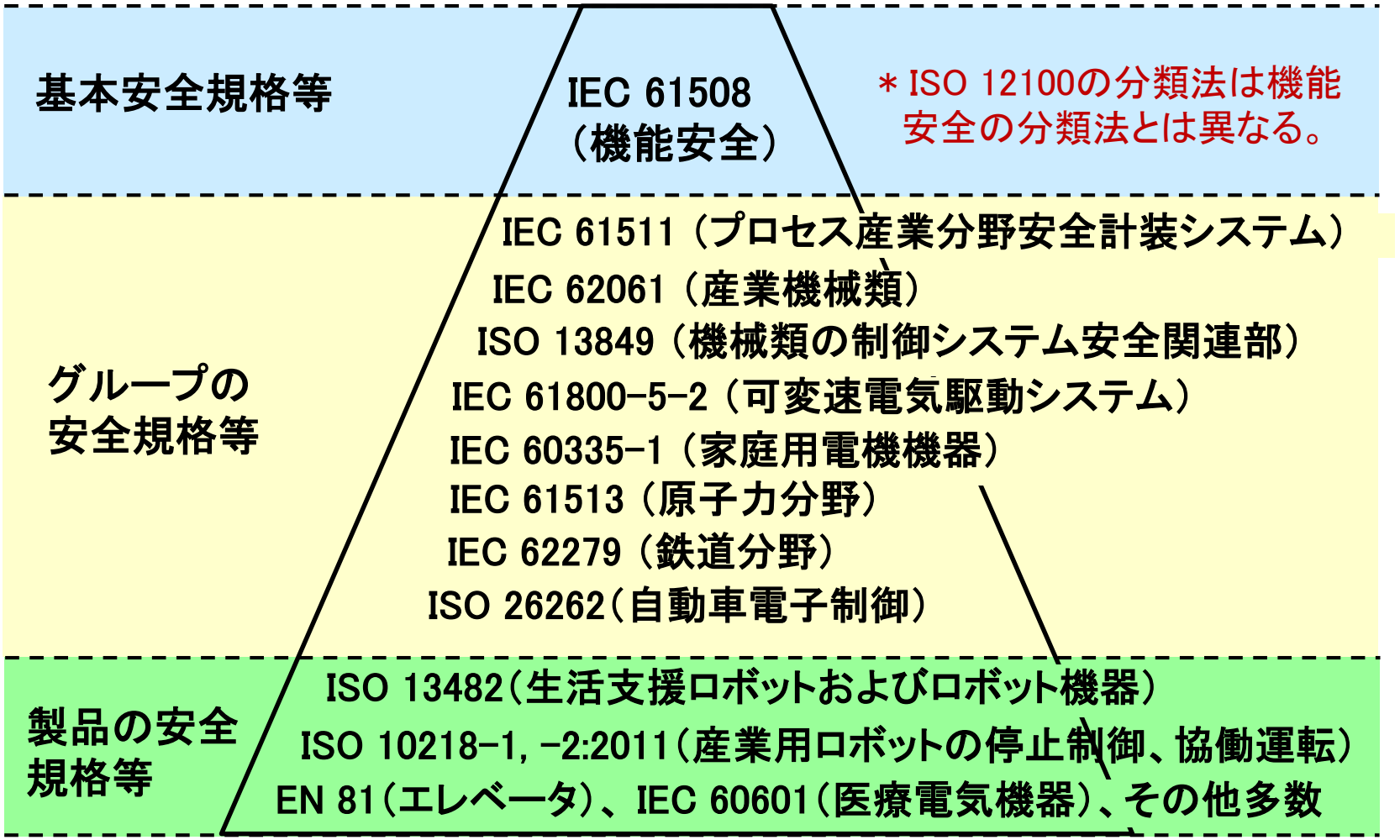
SDS：
Self-Driving System
(自動走行システム)
ESDS:
Electronic SDS
(電子自動走行システム)

電気・電子・プログラマブル電子安全関連系：
所与の安全機能を所与の安全度水準（SIL）で履行するシステム
安全機能：
全体システムの安全な状態を維持し、又は安全な状態に移行させる電気・電子・プログラマブル電子安全関連系の機能

出典：櫛引、佐藤、自動車技会論文集、Vol.41、No.1、pp.13-18、Jan.2010

機能安全関連規格類の現況

分類は、ISO/IECガイド51及び IEC ガイド104に従う*。



[出典：佐藤、機能安全／機械安全規格の基礎とリスクアセスメントーSIL, PL, 自動車用SILの評価法、日刊工業新聞社、2011]

電気･電子･プログラマブル電子安全関連系の機能安全
ー安全機能の不具合を起こす原因ー

- 知識上想定内かつ対策の想定外の原因 ー ハザード･リスク分析で見出した不具合原因
  - ハードウェアの故障
  - ヒューマンエラー
  - サイバー攻撃
  - その他予測可能な不具合原因

- 知識上かつ対策上の想定外原因 ー ハザード･リスク分析で見落とした不具合原因
  - ソフトウェアのバグ
  - ハザード同定で見落とした不具合原因
  - リスク分析上で見落とした不具合原因
  - 機能安全マネジメント実施上（検証、 妥当性確認、機能安全評価、安全要求仕様等）の不具合原因
  - 製品製造、検査、品質などに起因する不具合原因
  - その他予測不能な不具合原因

偶発的不具合原因に係る機能安全性能 - 安全度水準SIL
（既知の残留・残存リスク(対策の想定外既知事象)対応）

機能安全規格における安全度水準(SIL)

SIL：電気・電子・プログラマブル電子安全関連系に割り当てられる安全機能に対する偶発的原因に係る目標機能失敗確率

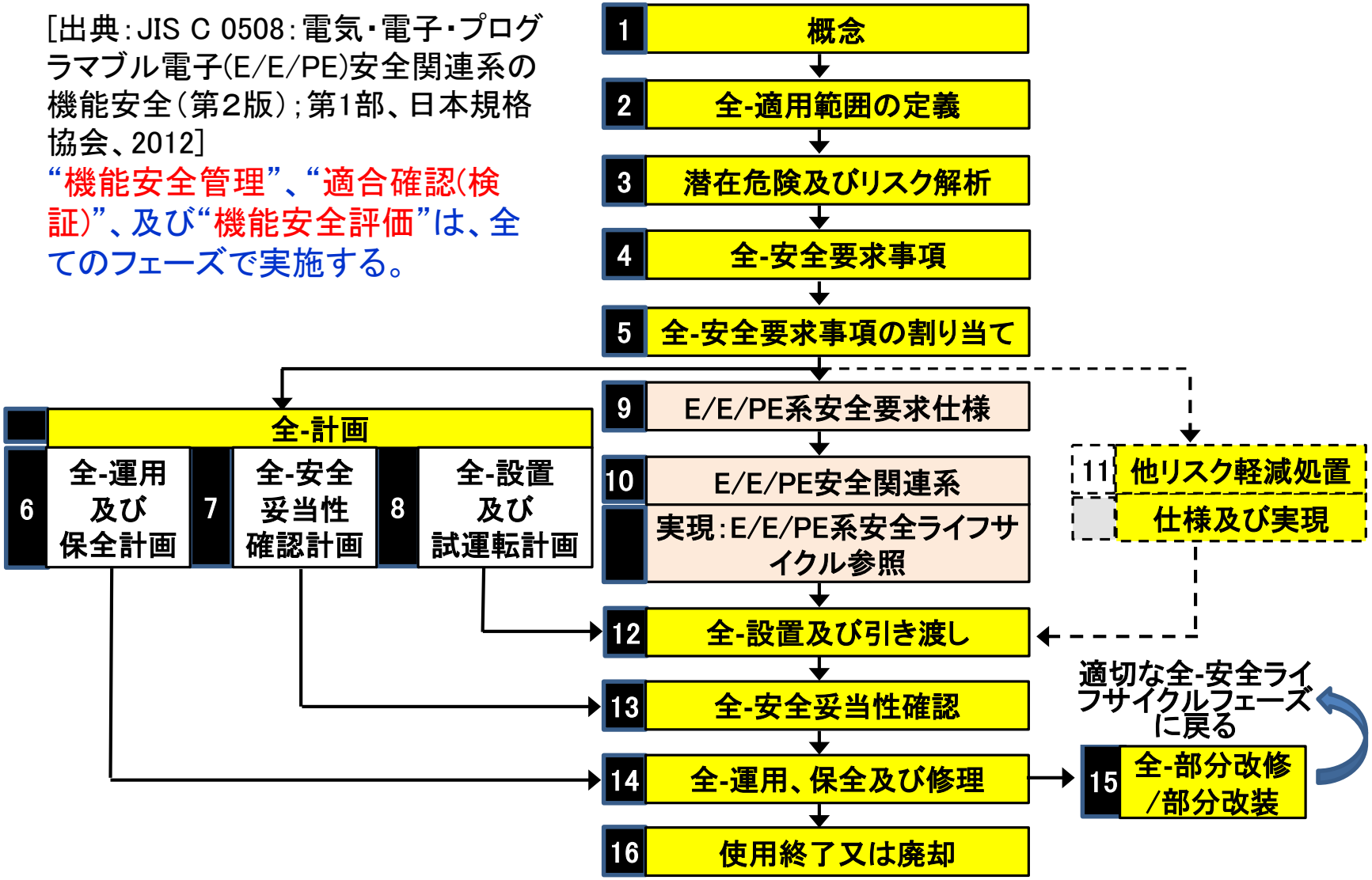
SIL	低頻度作動要求モード運用	高頻度作動要求または連続モード運用[1/時間]
4	10^{-5}以上10^{-4}未満	10^{-9}以上10^{-8}未満
3	10^{-4}以上10^{-3}未満	10^{-8}以上10^{-7}未満
2	10^{-3}以上10^{-2}未満	10^{-7}以上10^{-6}未満
1	10^{-2}以上10^{-1}未満	10^{-6}以上10^{-5}未満

[出典：JIS C 0508；第1部、日本規格協会、2012]

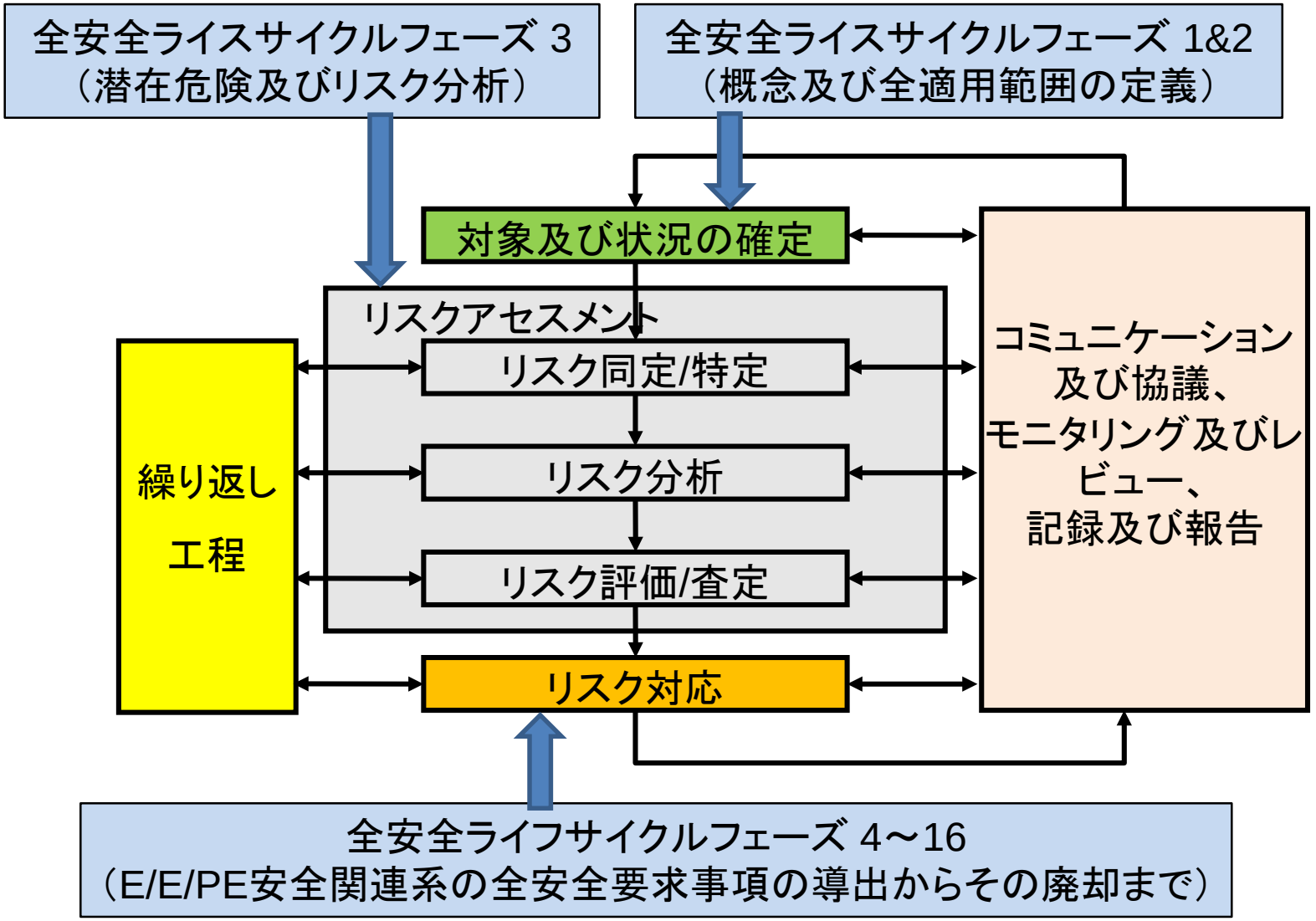
基本安全規格IEC 61508における全安全ライフサイクル
（危険なメタリスク(知識上かつ対策上の想定外事象)対応）

[出典：JIS C 0508：電気・電子・プログラマブル電子(E/E/PE)安全関連系の機能安全（第2版）；第1部、日本規格協会、2012]

“機能安全管理”、“適合確認(検証)”、及び“機能安全評価”は、全てのフェーズで実施する。



リスクマネジメントと機能安全とのプロセス整合



ソフトウェア（S/W）の安全性ー安全なS/Wとは？

安全性に影響する以下の特性を満たすものが安全なS/Wの必要条件である。

- S/Wに係る安全要求仕様に関する:（詳細省略）
- S/Wの設計・開発ー構造設計に関する:（詳細省略）
- S/Wの設計・開発ー支援ツール/プログラム言語に関する（詳細省略）
- S/Wの設計・開発ー詳細設計に関する:（詳細省略）
- S/Wの設計・開発ーモジュールテスト及び統合に関する:（詳細省略）
- S/Wのプログラマブル電子機器との統合に関する:(詳細省略）
- システムの安全妥当性確認におけるS/Wの位置づけ:（詳細省略）
- S/Wの部分改修(modification)に関する:
  - 要求事項に関する完全性
  - 要求事項に関する正確性
  - 固有設計フォールト導入がないこと
  - 望ましくない挙動がないこと
  - 検証可能かつテスト可能な設計
  - 回帰テスト(regression testing)及び検証の網羅性
- S/Wの検証に関する:（詳細省略）
- 機能安全評価に関する:（詳細省略）

[IEC 61508-3:2010を参照]

人工知能(AI）S/Wの安全性ーAIの安全な進化とは？

次のような要求事項を満たすことが必要条件となる：

- 安全関連AIの進化（S/Wの部分改修）によって影響を受ける全てのS/Wモジュール、必要とされる検証及び再設計が特定されること
- 安全関連AIの進化過程が明確であること
- 安全関連AIの進化は、あらかじめ定めた枠内で実施し、当該枠内の進化によって影響を受ける潜在危険群、進化の具体及び進化の理由が明確化されていること
- 安全関連AIの進化に関して、それぞれ主要な部分/サブシステムの属性が適切であるかを検証すること（S/Wアーキテクチャ設計完了後）
- 安全関連AIの進化に関して、S/W設計仕様のそれぞれ主要な要素の属性が適切であるかを検証すること（S/Wシステム設計完了後）
- 安全関連AIの進化に関して、それぞれのS/Wモジュールの属性が適切であるかを検証すること（それぞれのS/Wモジュール設計完了後）
- その他－省略

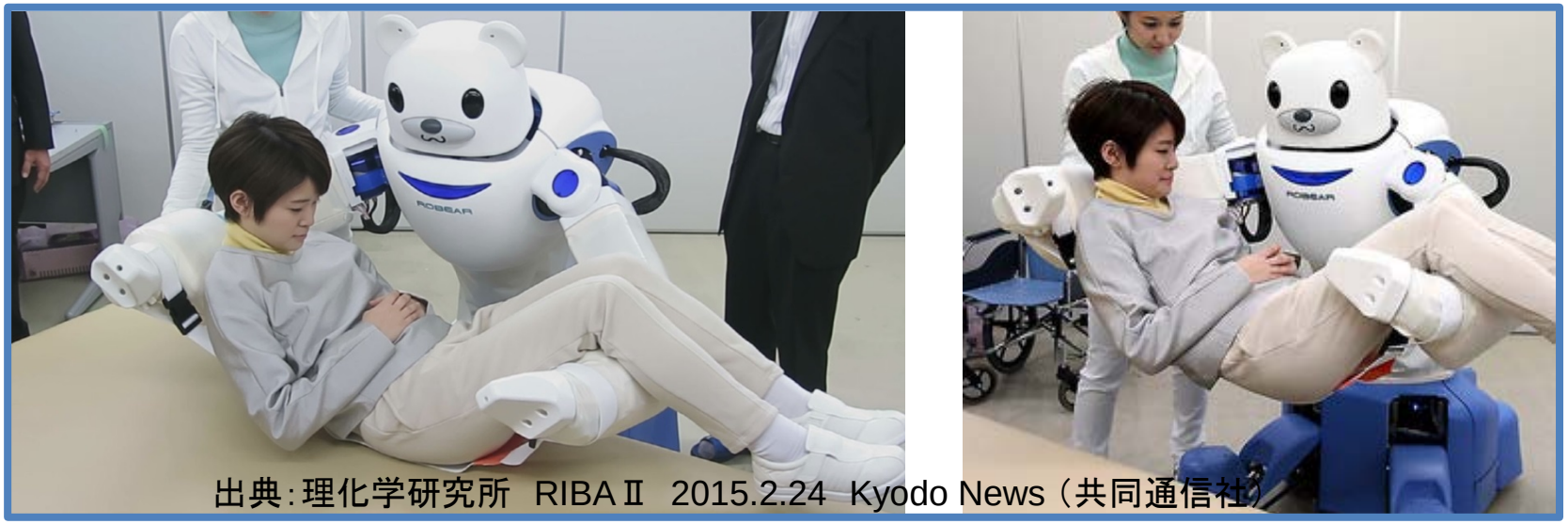
AI（を実装した）システムの安全性

- AI S/Wの安全確保

- 多重防護層によるシステム安全の確保（介護ロボット、自動運転車を事例として）

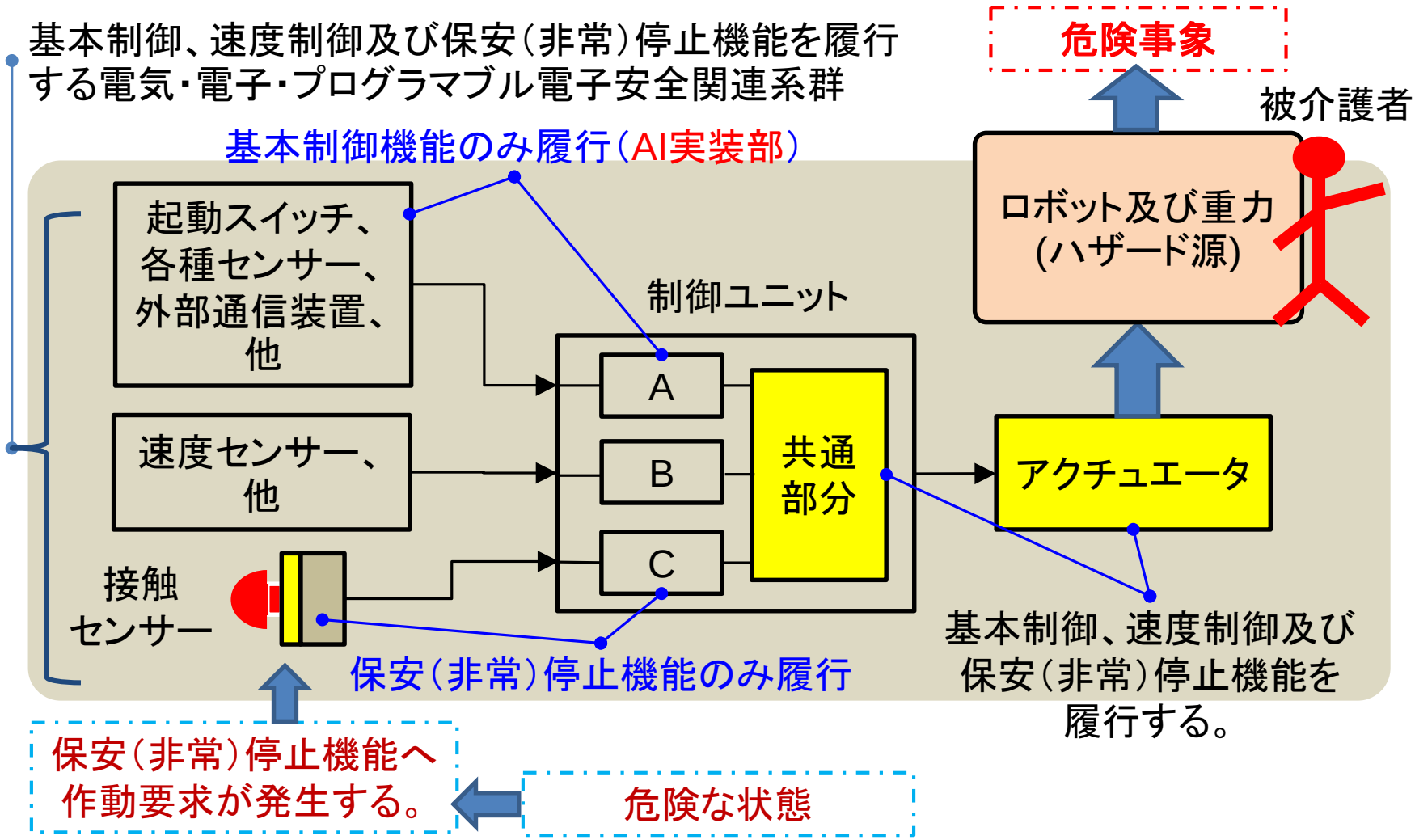
介護ロボットの多重防護層

介護ロボットのデモ

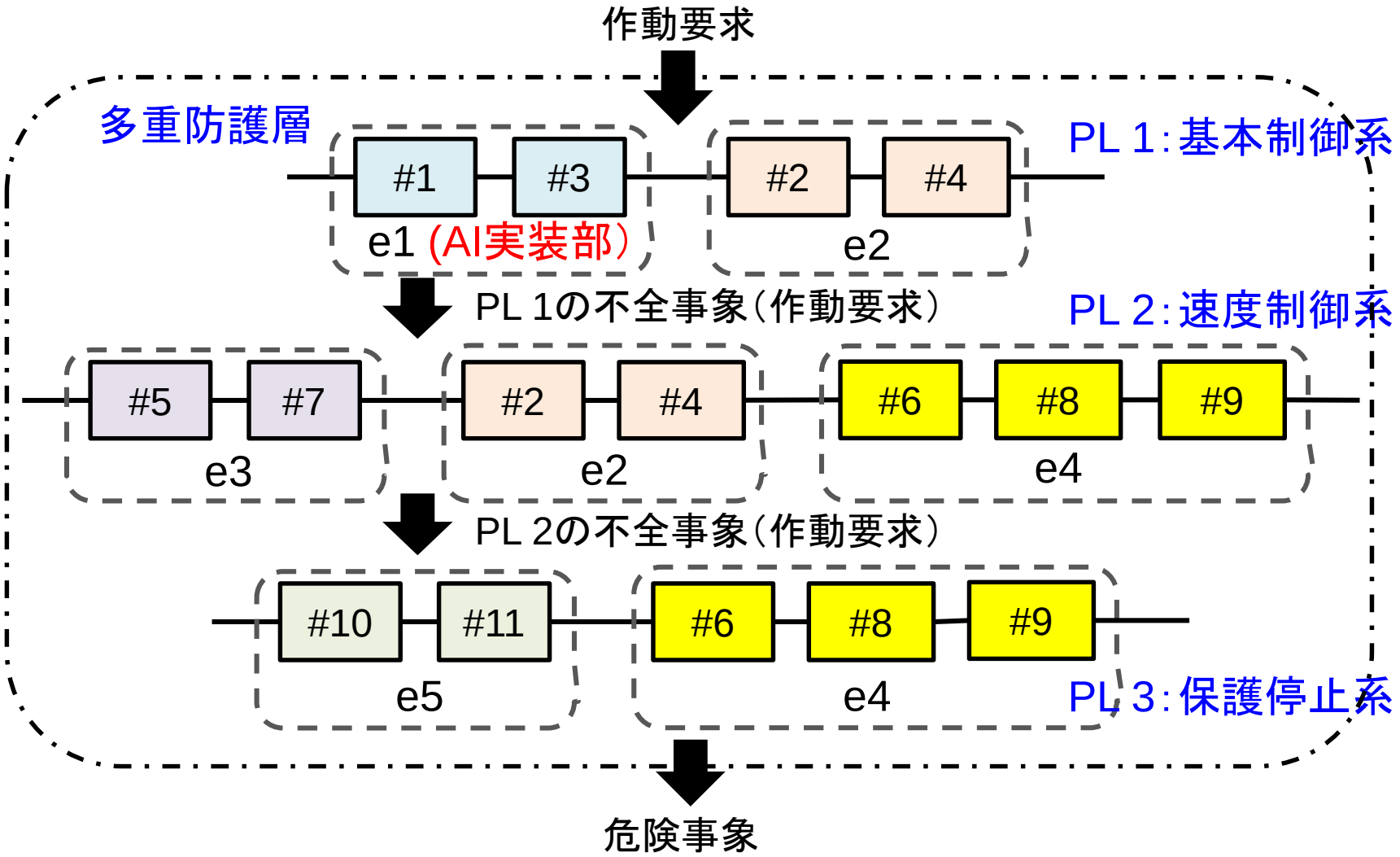


出典：理化学研究所　RIBAⅡ　2015.2.24　Kyodo News（共同通信社）

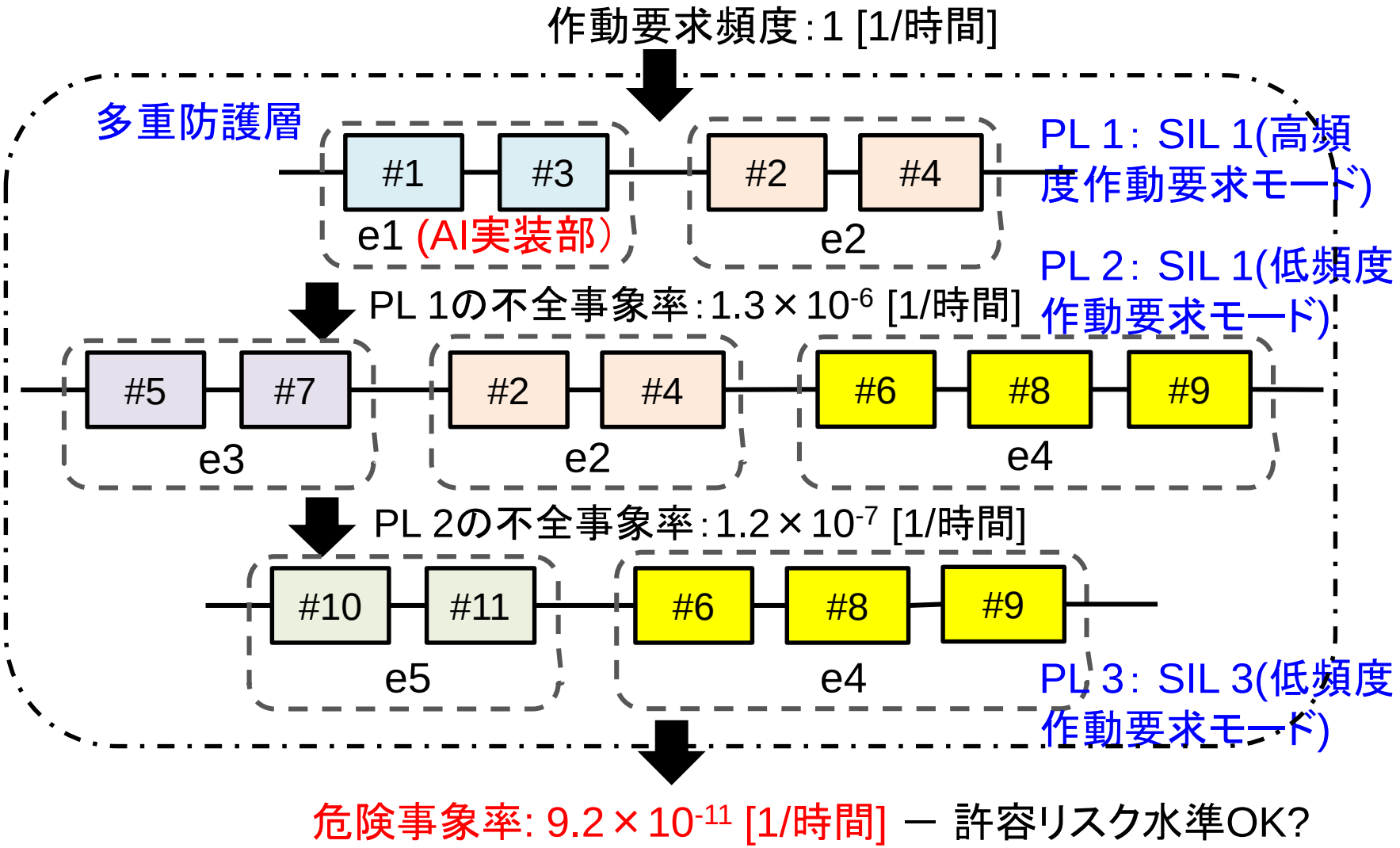
衝突ハザード及び転倒ハザード等をもつ介護ロボットの電気・電子・プログラマブル電子安全関連系群（多重防護層）



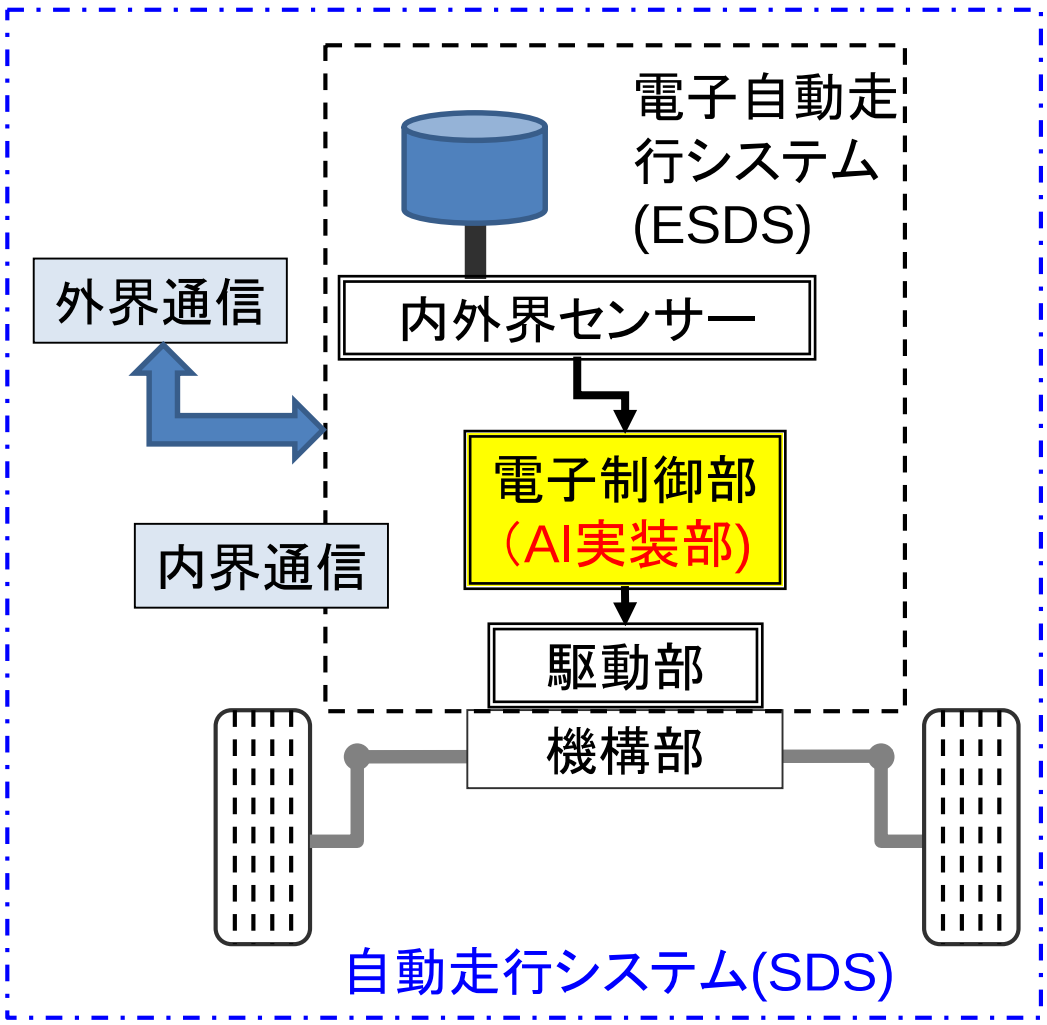
介護ロボットのAI実装多重防護層 – SILの割振り



介護ロボットのAI実装多重防護層のSILと残存リスク



自動運転における電子自動走行システム - 基本制御系

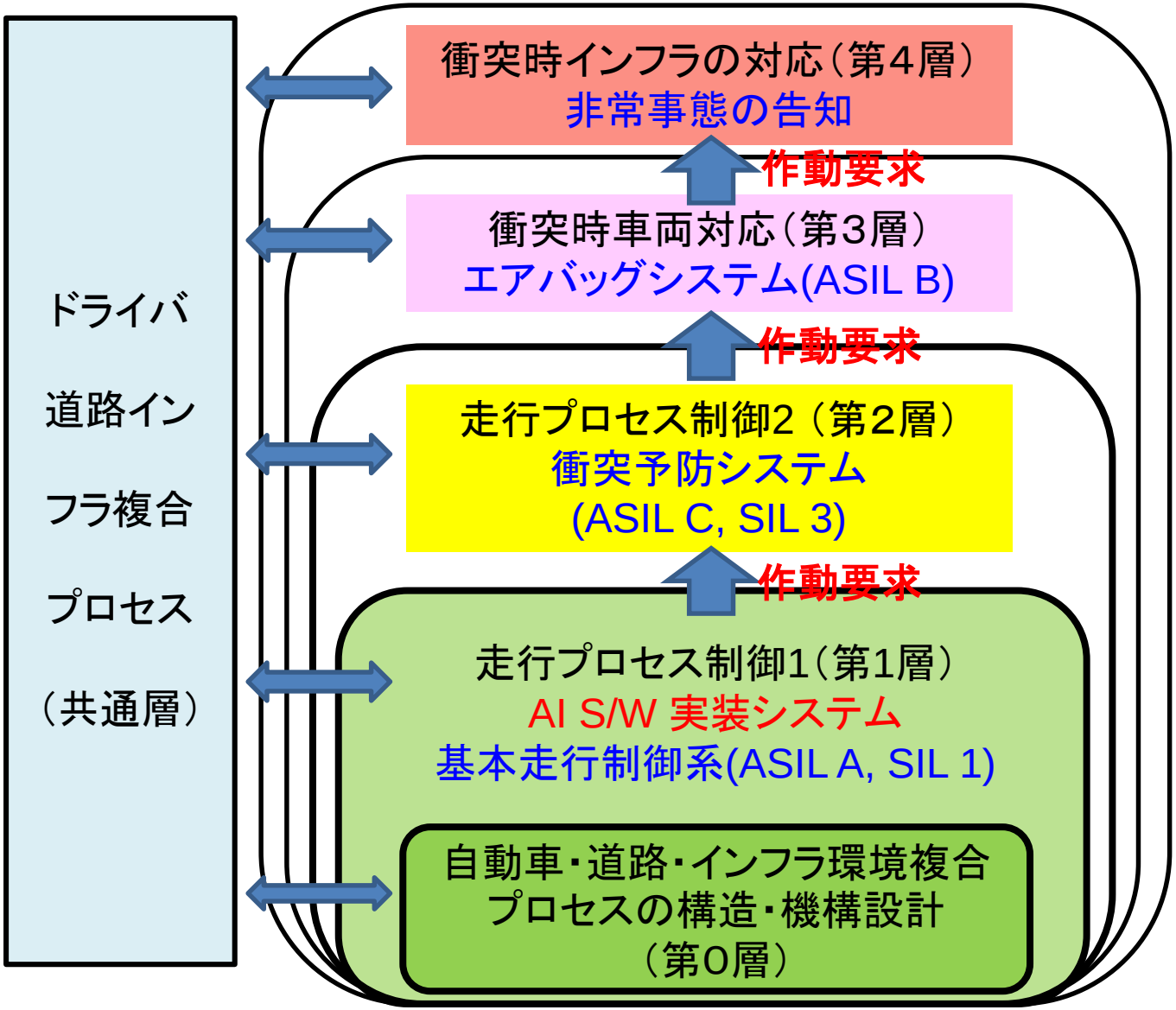


SDS :
Self-Driving System
(自動走行システム)

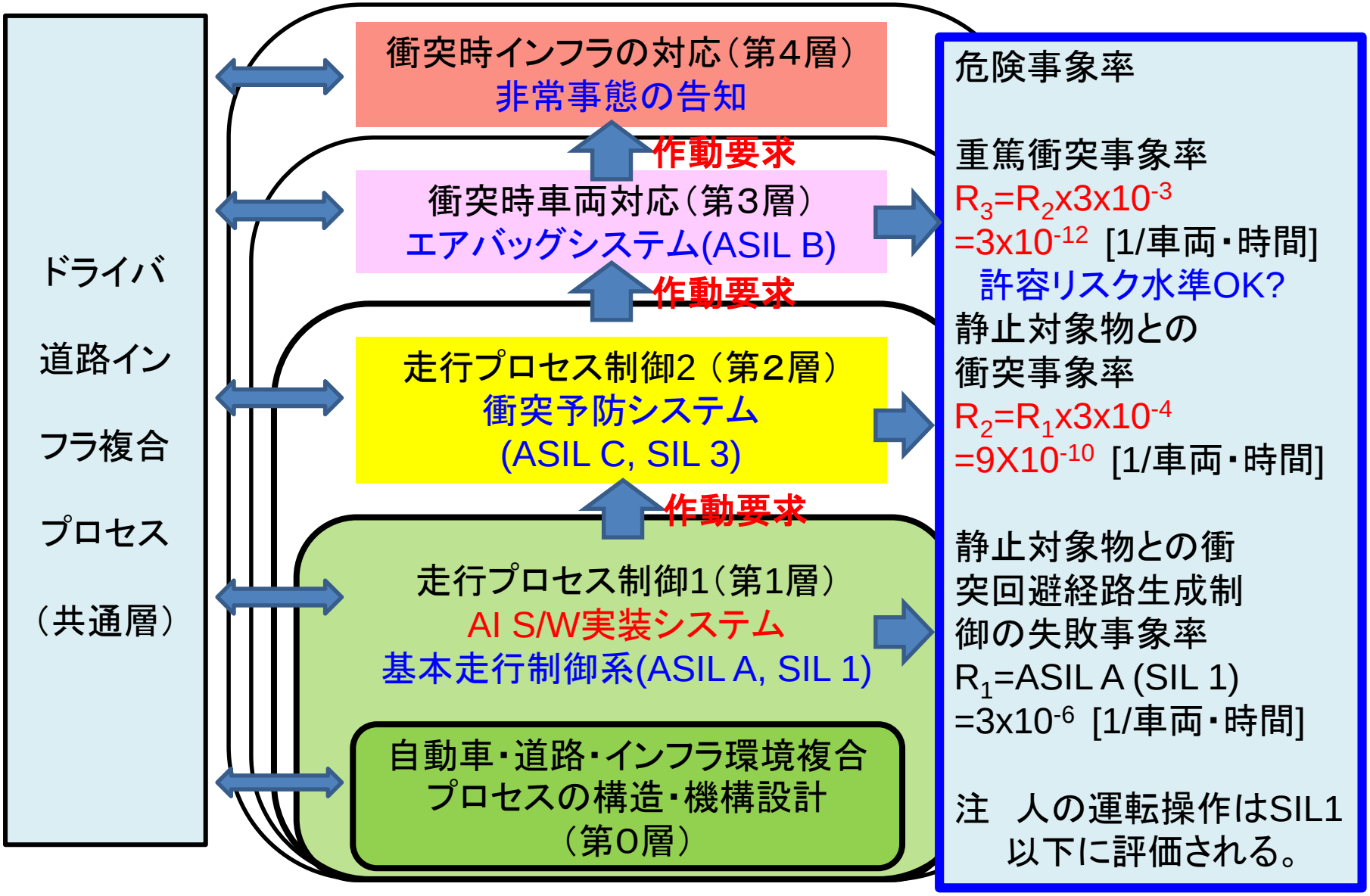
ESDS:
Electronic SDS
(電子自動走行システム)

出典:櫛引、佐藤、自動車技会論文集、Vol.41、No.1、pp.13-18、Jan.2010

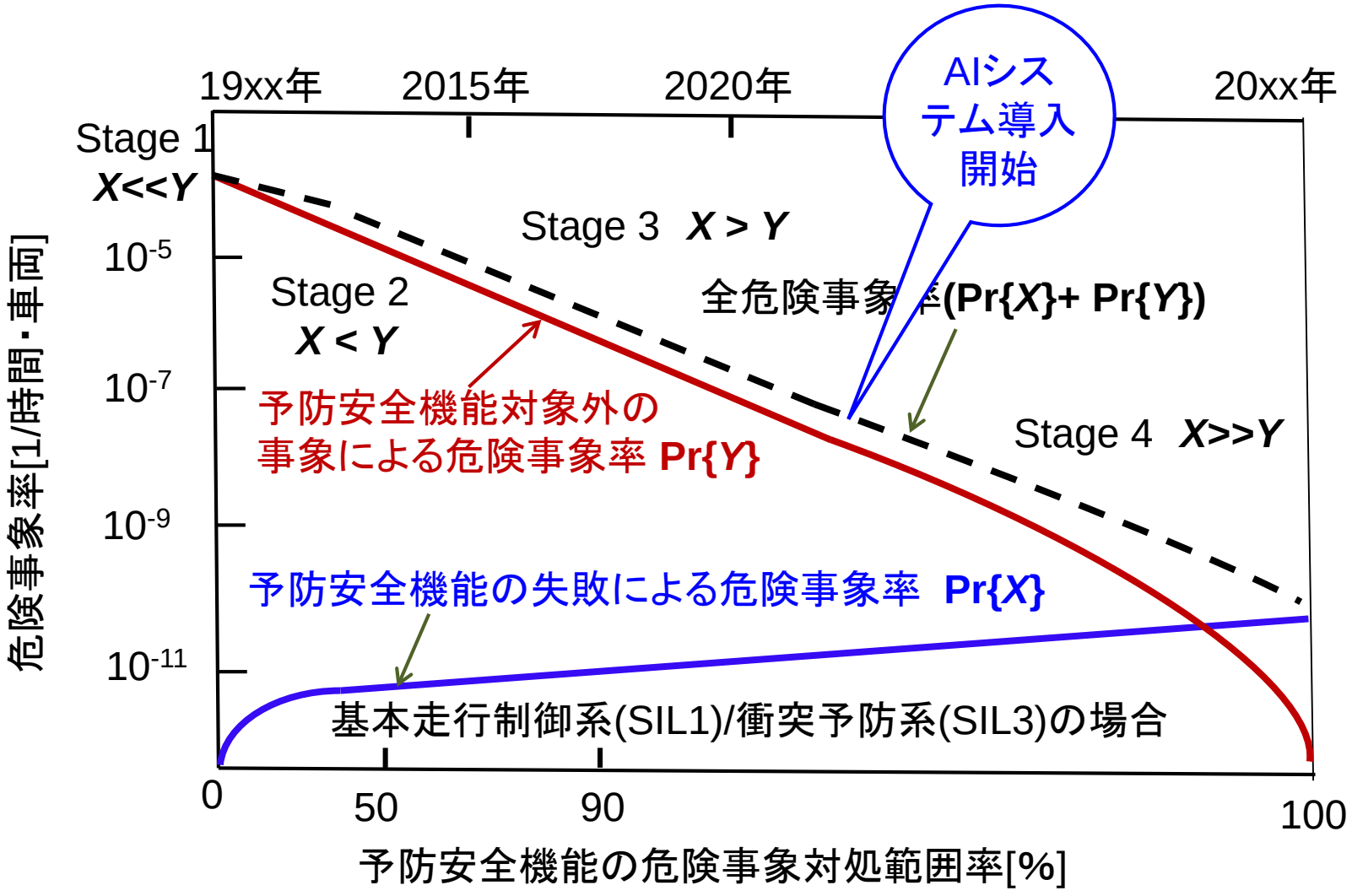
AI 実装自動運転車の多重防護層



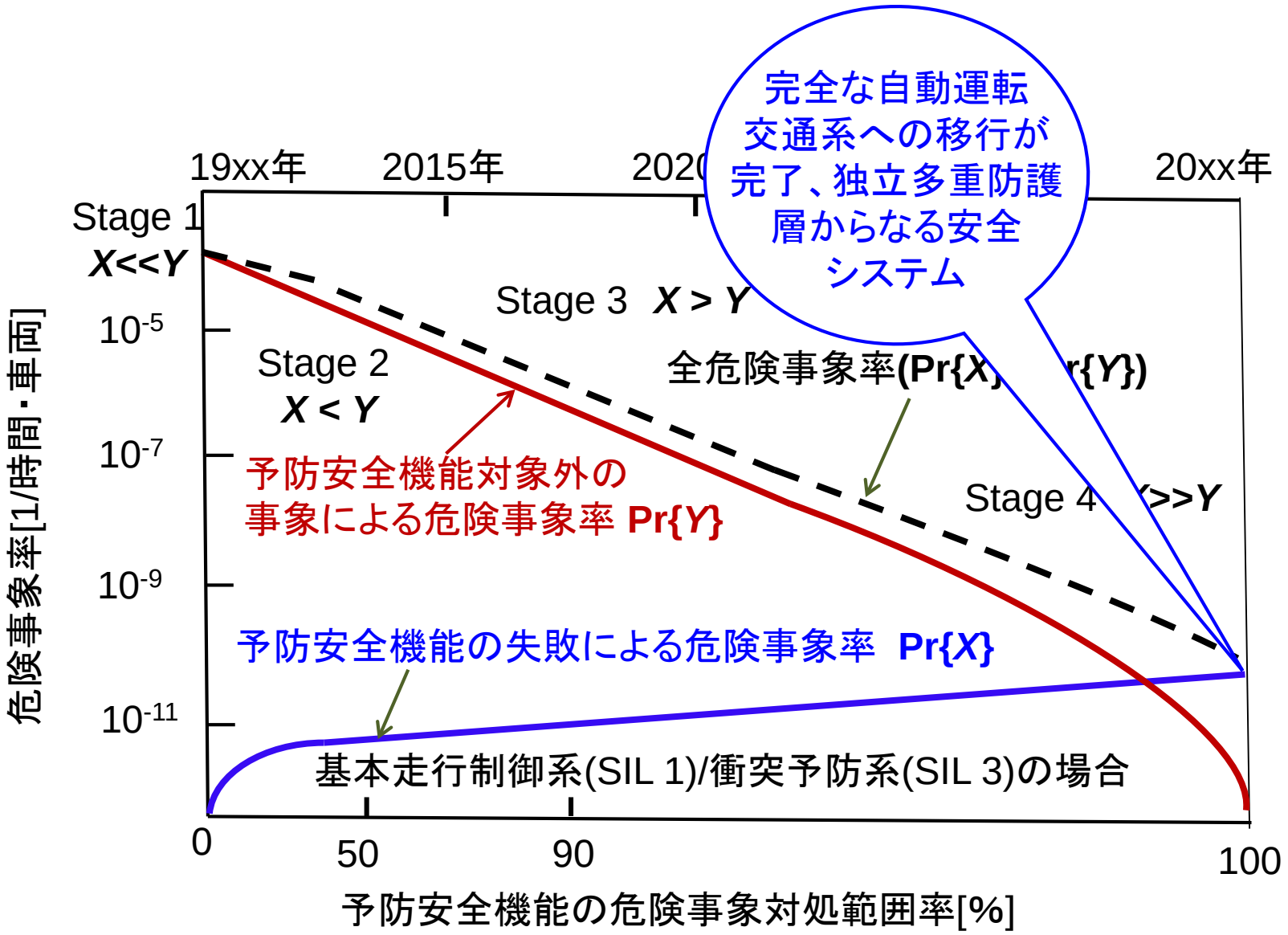
AI 実装自動運転車の多重防護層とリスク軽減/残存リスク



自動車予防安全と機能安全性能とによるリスク軽減目標



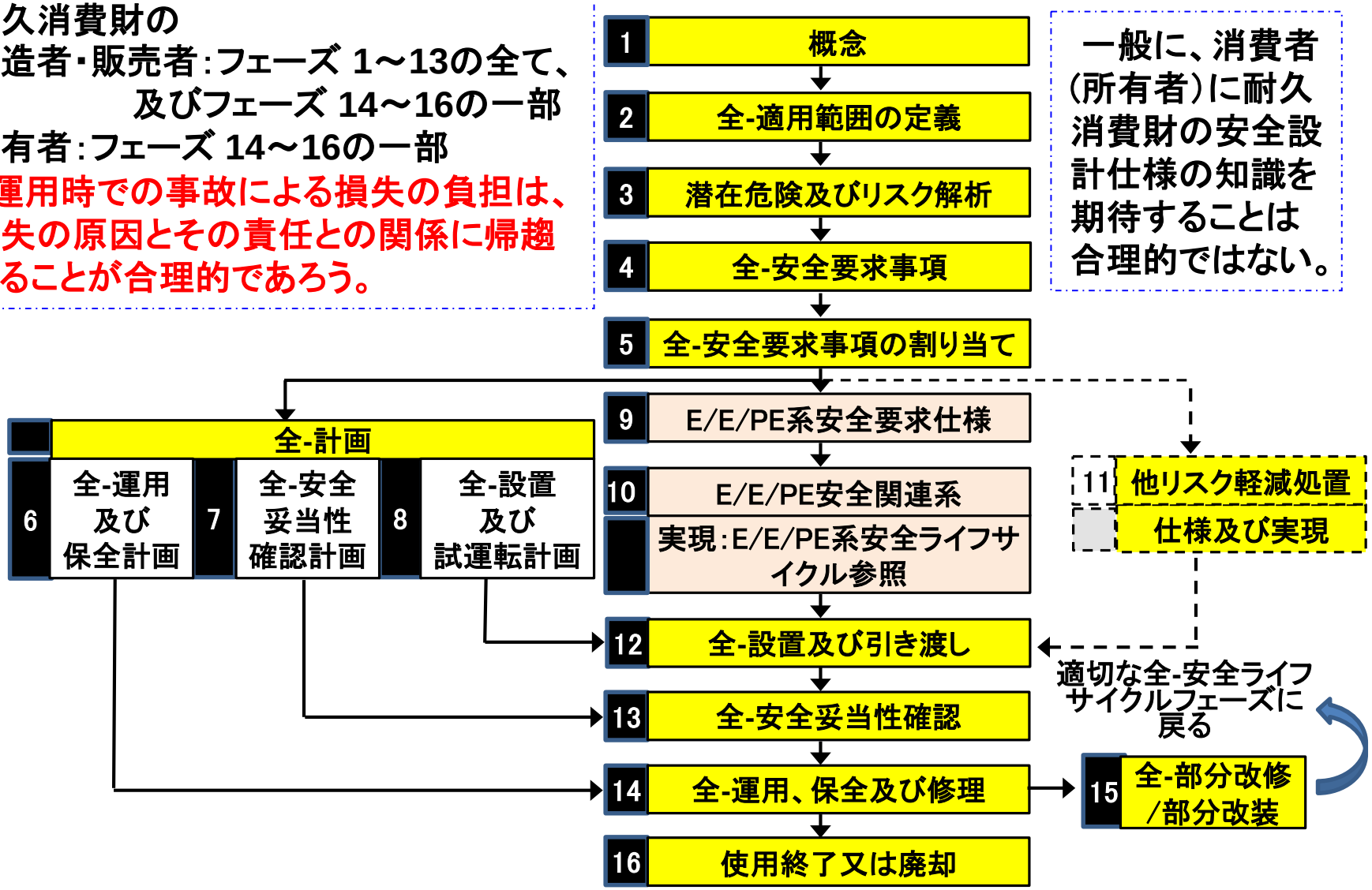
自動車予防安全と機能安全性能とによるリスク軽減目標



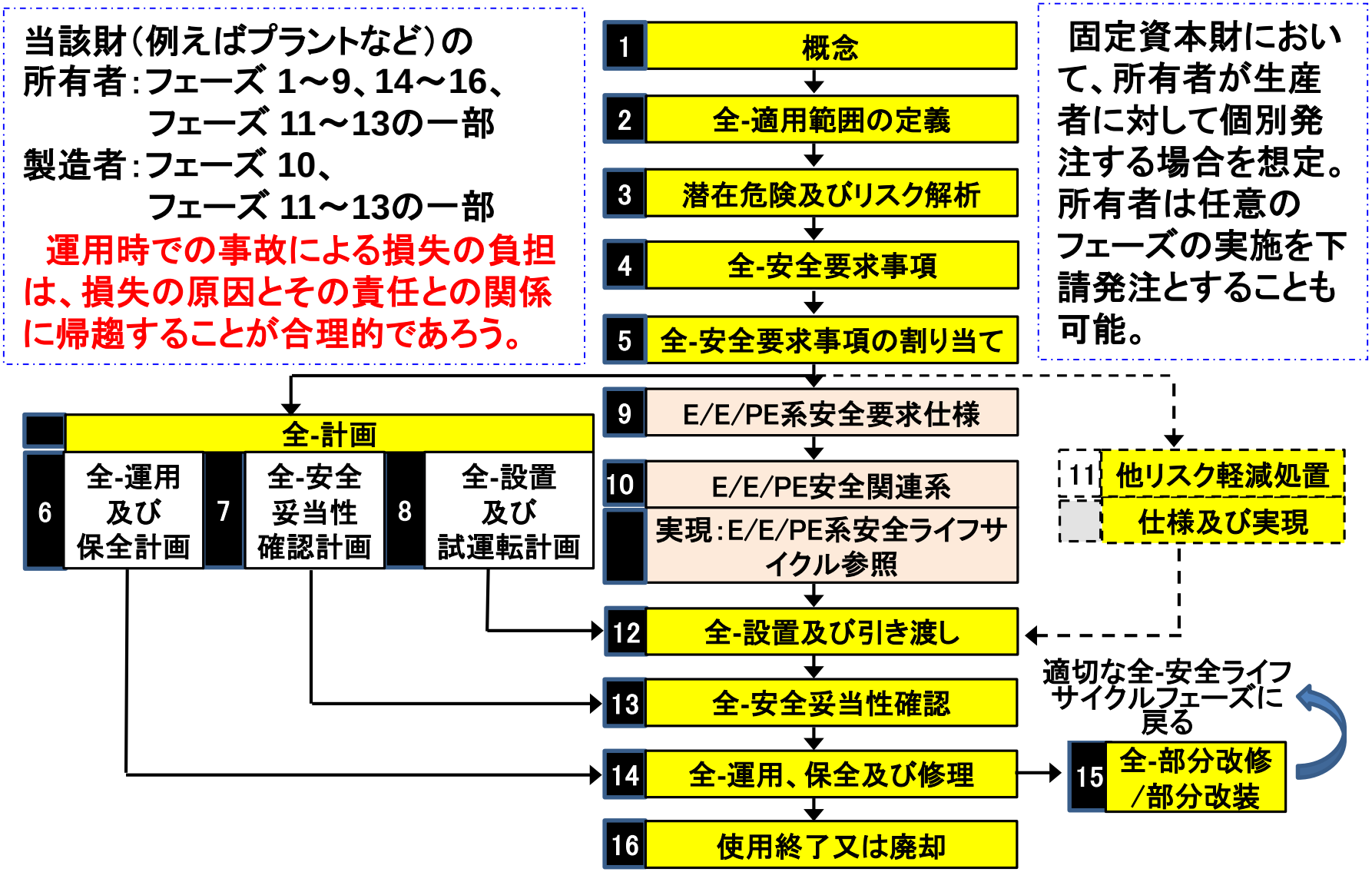
機能安全達成の責任と実施者 - 耐久消費財の場合

耐久消費財の
製造者・販売者：フェーズ 1～13の全て、
及びフェーズ 14～16の一部
所有者：フェーズ 14～16の一部
運用時での事故による損失の負担は、損失の原因とその責任との関係に帰趨することが合理的であろう。

一般に、消費者（所有者）に耐久消費財の安全設計仕様の知識を期待することは合理的ではない。



機能安全達成の責任と実施者 - 固定資本財の場合



まとめ

- AI S/Wを実装するシステムの安全確保は、
  - AIの安全確保による（方策A）、
  - AI以外の手段による(方策B）、
  - 方策A及びBによる（方策C）。
- 現在、機能安全基本規格IEC 61508の改訂作業が行われている。当該改訂作業で方策A、B、Cに対する機能安全確保に関する要求事項の補強・追補、及び指針を検討中である。
- 技術的に方策Aを達成あるいは保証することが困難な状況では、とりあえず方策BによってAI実装システムの安全確保を実施し、当該システムの運用経験によりAIの安全性を実証していくことも可能である。この場合、実証のための仕組みを全体システムに組み込む必要がある。
- 全体システムが機能安全規格に適合して設計・運用され、残存リスクがおおよそ定量的に把握されている場合、将来起こるかもしれない危害による損失及びその責任の分担の程度が想定可能である。これにより、想定される損失コストをあらかじめ製品/運用費用に含ませることも可能であろう。