ENAA　2023年度 第7回 エンジニアリングの最新ＤＸセミナー第3期
2024/1/26

AI法規制とAI安全規格の動向

三菱電機株式会社

神余　浩夫

MITSUBISHI ELECTRIC CORPORATION

0.1 はじめに

■ AI技術の健全な発展と市場拡大に向けて、各国はAI技術に関する規制と標準化に取り組んでいる。

- EU AI act：自動運転や協働ロボットなどハイリスク用途へのAI技術の適用。
- ISO/IEC TS 22440 AI - 機能安全とAIシステム-要求事項
  - ハイリスク用途（安全機能など）へのAI技術適用のための技術仕様・要求。
  - AI技術を安全関連部に適用する際の基本的な考え方、開発プロセス、技法や手段の推奨と制限など広い内容を含む。

■本講では、AI法規制とAI安全規格の動向について解説する

■対象者：AI製品・サービスの開発・品質保証に興味のある方

■目標：欧州AI法および関連規格について学ぶ

0.2 講師紹介

■神余　浩夫（かなまる　ひろお）

1987 大阪大学大学院工学研究科原子力工学修士課程修了

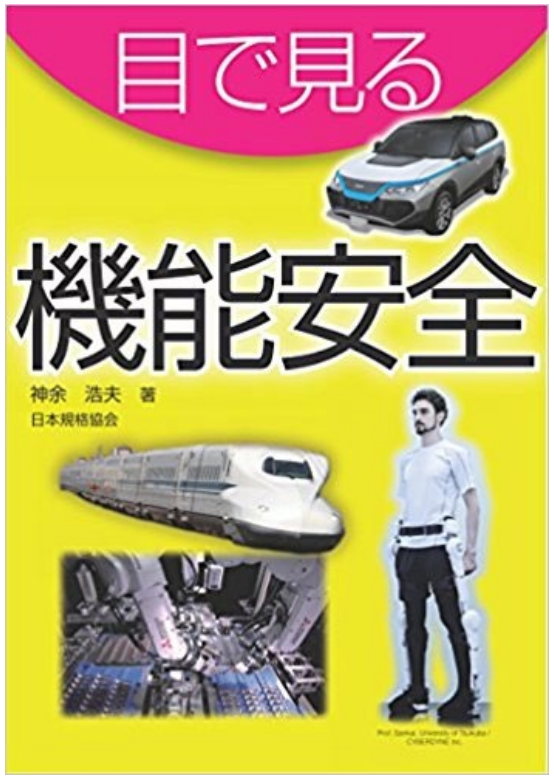
1987 三菱電機入社
中央研究所→産業システム研究所→先端技術総合研究所

2004 三菱電機名古屋製作所
安全シーケンサMELSEC Safety, CC-Link Safetyの開発

2011 三菱電機先端技術総合研究所
機能安全、制御セキュリティの研究・開発・技術指導
機械・ロボット安全の啓蒙・教育(経産省，厚労省)
TUV Rheinland 機能安全エキスパート(FSexp)
IEC 61508, IEC 62443 、ISO/IEC TR 5469他国際エキスパート

2023 三菱電機知的財産センター標準化戦略室

著書：目で見る機能安全，日本規格協会，2017



0.3 目次

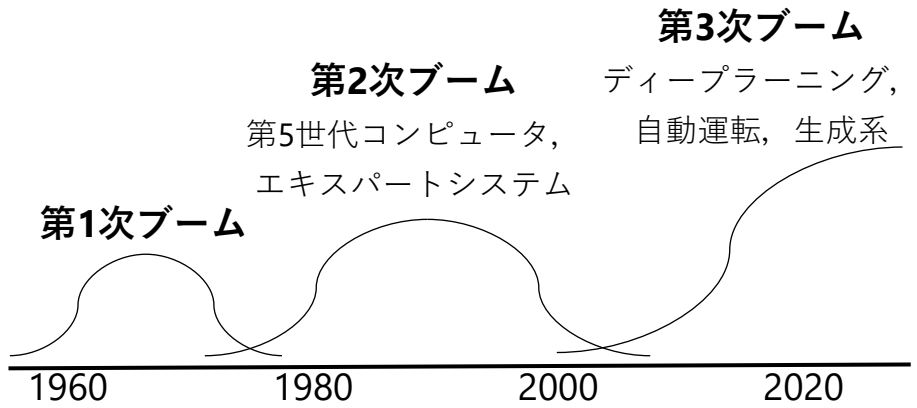
1 各国のAI規制，EU AI法

1.1 AIブームの歴史

■1960年代：第1次AIブーム(探索と推論)
- AIの登場(1956ダートマス会議)，自然言語処理
- ハード・ソフトウェアのパワー不足により衰退

■1980年代：第2次AIブーム(知識表現)
- 第5世代コンピュータ(ICOT)
  - MELCOM PSI – 40bitCPU/80MB、KL0/ESP
- エキスパートシステム＝知識ベース＋コツ・秘訣
- ハード・ソフトウェアのパワー不足により衰退

■2000年代：第3次AIブーム(機械学習)
- ビッグデータ、深層学習．自動運転、生成系、画像認識
- 各国の規制（AI倫理等）、標準化、認証制度
- ハード・ソフトウェアのパワー不足により衰退？



MELCOM PSI-コンピュータ博物館
(ipsj.or.jp)

Maisart

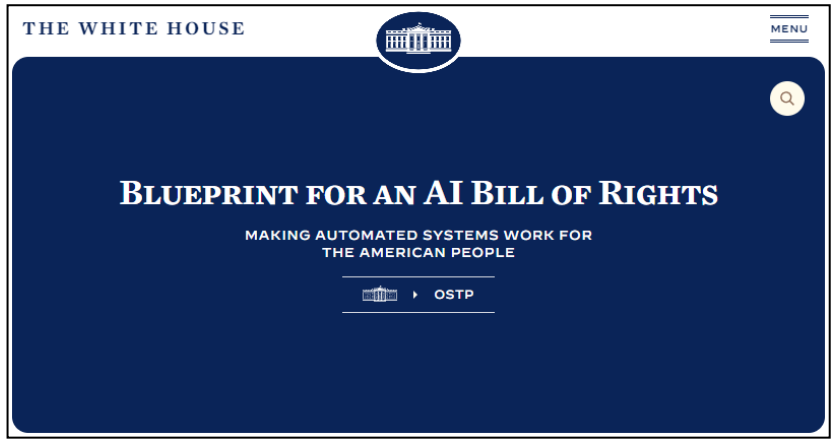
1.2 各国のAI安全に関する施策

■EU AI法 (AI Act)…現在審議中

- AIの安全・プライバシー等のリスクに対処し、開発投資を加速させるためのEU法令
  - 対象システムのリスクを、最小、限定的、高度、許容不可の4段階に分けて、高リスク用途の規制について言及。
  - 2023年12月9日，議会と理事会の合意→2027年発効?

■米国AI権利憲章(2022年10月)

- AI開発や仕様において考慮すべき５つの原則
  - 安全で効果的なシステム
  - アルゴリズムに基づく差別からの保護
  - データ・プライバシー保護
  - 使用者への通知と説明
  - 人間による代替、考慮、予備的措置
- 技術的側面よりも倫理性を重視
- 安全保障の側面



https://www.whitehouse.gov/ostp/ai-bill-of-rights/

1.3 各国のAI安全に関する施策

■日本 AIガバナンス

- Society5.0でもAI活用が推奨されており、人がAI技術に振り回されないためのAIガバナンスの必要性に迫られている。
  - 細かな行為義務を示すルールベースから、達成されるべき価値を示すゴールベースとし、非拘束的なガイドラインや標準を参照する方針をとる。
  - AI原則と企業の取組のギャップを埋める中間的なガイドラインが求められている。
  - 消費者側もAIと賢く向き合うためのリテラシー向上が必要。
  - 報告書は、中間的なガイドラインを利用するインセンティブの確保、他国との政策連携や標準化協力、モニタリングとエンフォースメントを課題として挙げている。

1.4 最近の各国の論調（内閣府「AIを巡る主な論点」-2023年5月11日AI戦略会議資料）

- 米国行政管理予算局(OMB)は、国民の権利等の保護のため、政府機関におけるAI利用についてガイダンスを公開し、意見募集を行うと発表。
- ホワイトハウスは、新たに7つの国立AI研究機関を立ち上げるため、1億4000万ドルの資金提供を発表。気候、農業、エネルギー、公衆衛生、教育、サイバーセキュリティ等の重要分野における取組を促進。

- プライバシー・個人情報保護法(PIPEDA)の下、政府がプライバシーに関する懸念点を調査中。

- 競争・市場庁(CMA)が、基盤モデルの開発と利用における競争確保と消費者保護についての調査を開始。（注）各種報道資料などから内閣府作成。
- AI開発向け等の大規模計算資源の整備に約９億ポンドを投資。また、今後10年間、AIに関する優れた研究に対し、毎年100万ポンドの賞金を授与することを決定。

- データ保護当局(Garante)が、利用者の年齢確認や情報提供義務、法的根拠を特定できていない点、正確性原則違反などを理由に一時的にChatGPTの利用を禁止。その後、OpenAIが対応措置を講じたことから禁止を解除。

- データ保護当局(CNIL)は、ChatGPTに対する複数の申し立てに基づき調査を実施中。

- EU加盟国のデータ保護当局等が構成する「欧州データ保護会議」(EDPB)がChatGPTを取り扱うタスクフォースを設置。各データ保護当局の協力と情報共有を目的としているが、AIに関する包括的なプライバシーポリシーの確立に向かうのではとの見方もあり。

- サイバー空間管理機関(CAC)が、生成AIに関して、公衆向けサービスの提供前に当局に対して安全性評価を提出すること、生成AIの出力は共産主義の基本的な価値観に沿うものとすべきこと等を求める規制案を公表。

- 政府主導プログラムの下で、インド独自の生成AI「BharatGPT」を開発中。23の公用語と6000の方言があると言われるインドで重要な異なる言語間の翻訳・コミュニケーションを主眼に、独自のデータセットを用いてLLMを開発している。

1.5 EU AI法

■AIに関する初の規制法

- 2021 年 4 月、欧州委員会は AI に関する初の EU 規制枠組みを提案した。さまざまな用途で利用できるAIシステムを分析し、ユーザーにもたらすリスクに応じて分類している。リスクレベルが異なると、多かれ少なかれ規制が必要になる。承認されれば、これらはAIに関する世界初のルールとなる。

● 議会が AI 法案に求めていること

- 議会の優先事項は、EU で使用される AI システムが安全で、透明性があり、追跡可能で、差別がなく、環境に優しいものであることを確認する。AI システムは、有害な結果を防ぐために、自動化ではなく人間によって監視される必要がある。
- 将来の AI システムに適用できる、テクノロジーに中立な AI の統一定義を確立したい。

● 主題(AI法第1条)

(a1) EU におけるAI システムの市場投入、運用および使用に関する調和された規則
(a2) 特定の人工知能の実践の禁止
(b) 高リスク AI システムに対する特定の要件とオペレーターに対する義務
(c) 特定の AI システムに対する調和された透明性ルール
(d) 市場監視、市場監視およびガバナンスに関する規則
(e) イノベーションを支援する措置

https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

1.6 適用対象とペナルティ

■域外適用→日本にも適用される

- AIシステム・サービスを提供すれば日本にも適用される（2条1項(a)号※）
- 正確には、EUにおいてAIシステムを市場に置き又はサービスを提供した提供者（など）に適用

■適用対象となるAIシステム

● 次の二つの要件を満たすもの（3条1項）

- 1.付属書Iの技法及びアプローチで開発されたソフトウェア
  - ディープラーニングを含む様々な方法を用いた……機械学習によるアプローチ
  - 知識表現、帰納（論理）プログラミング……を含む論理ベース及び知識ベースのアプローチ
  - 「統計的アプローチ」など
- 2.人間が定めた一定の一連の目的のために、当該ソフトウェアが相互作用する環境に影響を与えるコンテンツ、予測、推奨又は決定などのアウトプットを生成することができるもの

● 違反へのペナルティ（制裁金）

- 最大で3,000万ユーロ（約40億円）か全世界売上高の6%のうちどちらか高い金額（71条）
- 一定の違反や、基本権にリスクが生じ得る場合などに、適切な対応をしないと、AIシステムの市場からの取下げやリコールなどの是正措置を公的な機関から義務付けられる可能性がある（65～68条）

1.7 AI法：リスクレベルごとに異なるルール

・AIによるリスクのレベルに応じて、プロバイダーとユーザーの義務が定められている。

リスクレベル	利用	対象	
非許容リスク	禁止	EUの価値観と矛盾するAIの禁止 ・潜在意識への操作 ・一般人の分類を行う生体分類システム ・社会的スコアの一般的な利用 ・公的空間での法執行目的の遠隔生体認証 ・犯罪行為の発生，再発予防などへの利用	・不特定多数を対象とした顔認識データベース ・公的機関における人の感情を推測 ・公共空間の録画画像の事後分析
ハイリスク	要件と適合性評価の準拠	規制対象製品の安全要素 ・産業機械，医療機器など第三者認証の対象 特定分野および健康・安全・基本的権利 ・生体イ及び生体に基づくシステム ・重要インフラの管理と運用 ・教育と職業訓練 ・雇用，労働者管理	・環境に重大なリスク ・必須の民間サービスや公共サービス ・法による強制 ・移住，亡命および国境管理 ・司法運営と民主的プロセス
限定リスク	情報/透明性の義務	透明性義務が適用 ・一般人と相互作用する ・感情推定や生体情報に基づく分類 ・ディープフェイク	
最小リスク	制限なし	上記以外のAIシステム	

1.8 EUのAI規制（AI act）

リスクレベル	要求事項		
非許容リスク	－		
ハイリスク	・リスク管理プロセスの確立 ・高品質な学習，検証，テストデータの利用 ・文書化，ログ機能 ・適切な透明性確保，情報提供 ・人間による監視 ・堅牢性，正確性，セキュリティ ・ガイドラインや法令の考慮	AI提供者の義務 ・透明性，説明可能性 ・品質マネジメントシステム ・技術文書の作成 ・ログ記録6か月保管 ・適合性評価と変更時の再評価 ・EUデータベースへの登録 ・CEマーキング貼付，適合宣言書 ・市場モニタリング ・市場監視当局への協力 ・アクセシビリティ要件	AI利用者の義務 ・取扱説明書の遵守 ・人間によるAIシステムの監視 ・リスクに対する監視 ・重大事故・誤動作の通知義務 ・既存の法的義務の遵守
限定リスク	・人とAIシステムの相互作用と人への通知 ・感情認識または生体認証が使われていることの人への通知 ・ディープフェイクの警告ラベルを添付		
最小リスク	・必須義務はない		

具体的な技術要件は，整合規格(ISO/IECまたはEN規格)を参照

1.9 標準化と認証スキーム

■標準化（整合規格）

- 法令では具体的な技術基準まで言及できない、技術の進歩に追いつけない。
  →国際標準やEN規格を参照
  - ISO/IEC/JTC1/SC41 AI
  - CEN-CENELEC/JTC21 AI …EN規格の策定

■認証スキーム

- 製品・サービスの規格適合性評価
- ETSIが開発中(Oxford大のCAPを利用)
  - ETSI: European Telecommunications Standards Institute
  - IECEE/CMCが開発検討開始
  - TUV等に聞いたところ，審査できる人がいない


Text

EUROPEAN COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final
2021/0106(COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

EXPLANATORY MEMORANDUM

1. CONTEXT OF THE PROPOSAL

1.1. Reasons for and objectives of the proposal

2 ISO/IEC/JTC1/SC42 Artificial Intelligence

2.1 AIの標準化団体

■ISO/IEC JTC1 Information technology（情報技術）

- ISO/IEC JTC 1とは、ISOとIEC の合同技術委員会 (Joint Technical Committee)のこと。情報技術分野の標準化を行うための組織である。1987年設立。
  - 標準3461本発行、502本開発中、参加国39（オブザーバ63か国）
  - リエゾン　EC, Ecma International, ITU, IEEE, SBS(Small Business Standards)

■ISO/IEC JTC1/SC42 Artificial Intelligence（人工知能）

- 人工知能に関するJTC 1の標準化プログラムの中心となり、推進役を務める
- 人工知能アプリケーションを開発するJTC 1、IEC、ISO委員会へのガイダンスの提供
  - 2017年設立．標準24本発行、31本開発中

■CENELEC/JTC21 – AI

- EN規格-EU AI法の整合規格を担当

JTC1/SC42 Tokyo meeting (2019, 産総研お台場)

2.2 ISO/IEC JTC1/SC42 - AI

■ 組織構成

WG1 Foundational standards – 基礎規格
WG2 Data - データ
WG3 Trustworthiness – 信頼性
WG4 Use cases and applications -ユースケースとアプリケーション
WG5 Computational approaches and computational characteristics of AI systems - AIシステムの計算アプローチと計算特性
AHG4 Liaison with SC 27 (security) - SC27（セキュリティ）とのリエゾン
AHG7 JTC1 joint development review -共同開発レビュー
JWG2 Joint with JTC1/SC7: Testing of AI-based systems - AIベースシステムのテスト
JWG3 Joint with ISO/TC215: AI enabled health informatics - AIを活用した医療情報
JWG4 Joint with IEC/SC65A: Functional safety and AI systems - 機能安全とAIシステム
JWG5 Joint with ISO/TC37 WG: Natural language processing -自然言語処理

■ 日本の所轄団体

- 情報処理学会/情報規格調査会(ITSCJ)
- SC42専門委員会（議長：産総研　杉村領一氏）

一般社団法人　情報処理学会
情報規格調査会
Information Technology Standards Commission of Japan

2.3 SC42標準（発行済）

- TS 4213:2022 Assessment of machine learning (ML) classification performance
- IS 5338:2023 AI system life cycle processes
- IS 5339:2024 Guidance for AI applications
- TR 5469:2024 Functional safety and AI systems
- IS 8183:2023 Data life cycle framework
- IS 20546:2019 Big data - Overview and vocabulary
- TR 20547-x:20xx Big data reference architecture
- IS 22989:2022 AI concepts and terminology
- IS 23053:2022 Framework for AI - Systems Using Machine Learning (ML)
- IS 23894:2023 Guidance on risk management
- TR 24027:2021 Bias in AI systems and AI aided decision making
- TR 24028:2020 Overview of trustworthiness in artificial intelligence
- TR 24029-1:2021 Assessment of the robustness of neural networks - Part 1: Overview
- IS 24029-2:2023 Assessment of the robustness of neural networks - Part 2: Methodology for the use of formal methods
- TR 24030:2021 Use cases
- TR 24368:2022 Overview of ethical and societal concerns
- TR 24372:2021 Overview of computational approaches for AI systems
- IS 24668:2022 Process management framework for big data analytics
- IS 25059:2023 Software engineering - SQuaRE - Quality model for AI systems
- IS 38507:2022 Governance of IT - Governance implications of the use of artificial intelligence by organizations
- IS 42001:2023 Management system

2.4 SC42標準（開発中）

- DIS 5259-x: Data quality for analytics and ML
- FDIS 5392: Reference architecture of knowledge engineering
- CD TS 6254: Objectives and approaches for explainability and interpretability of ML models and AI systems
- DTS 8200: Controllability of automated AI systems
- DTS 12791: Treatment of unwanted bias in classification and regression machine learning tasks
- CD 12792: Transparency taxonomy of AI systems
- AWI TS 17847: Verification and validation analysis of AI systems
- DTR 17903: Overview of machine learning computing devices
- AWI TR 18988: Application of AI technologies in health informatics
- CD TR 20226: Environmental sustainability aspects of AI systems
- AWI TR 21221: Beneficial AI systems
- AWI TS 22440: Functional safety and AI systems - Requirements
- AWI TS 22443: Guidance on addressing societal concerns and ethical considerations
- AWI 23282: Evaluation methods for accurate natural language processing systems
- AWI 24029-3: Assessment of the robustness of neural networks — Part 3: Methodology for the use of statistical methods
- DTR 24030: AI Use cases
- TS 25058: Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of AI (AI) systems
- AWI TS 29119-11: Software and systems engineering — Software testing — Part 11: Testing of AI systems
- DIS 42005: AI system impact assessment
- DIS 42006: Requirements for bodies providing audit and certification of AI management systems
- AWI 42102: Taxonomy of AI system methods and capabilities
- AWI TR 42103: Overview of synthetic data in the context of AI systems
- AWI 42105: Guidance for human oversight of AI systems
- AWI TR 42106: Overview of differentiated benchmarking of AI system quality characteristics

3 ISO/IEC 42001 AI – Management system

3.1 ISO 42001 AIマネジメントシステム

■ISO/IEC 42001:2023 AI - Management system

■概要

- 人工知能（AI）を利用した製品やサービス（AIシステム）の普及が急速に進む中、AIシステムを安全・安心に利活用するためには、リスクベースアプローチ等を通じて、AIシステムを適切に開発・提供・使用することが必要です。
- 今回発行されたAIマネジメントシステムの国際規格により、AIに関するリスクを回避するための要件やリスクが生じた場合の対応を含む信頼性の高いマネジメントシステムが構築可能となり、より安全・安心なAIシステムの普及拡大への貢献が期待されます。
  - （METIニュースリリース2023年1月15日）

3.2 ISO 42001 AIマネジメントシステム

■背景

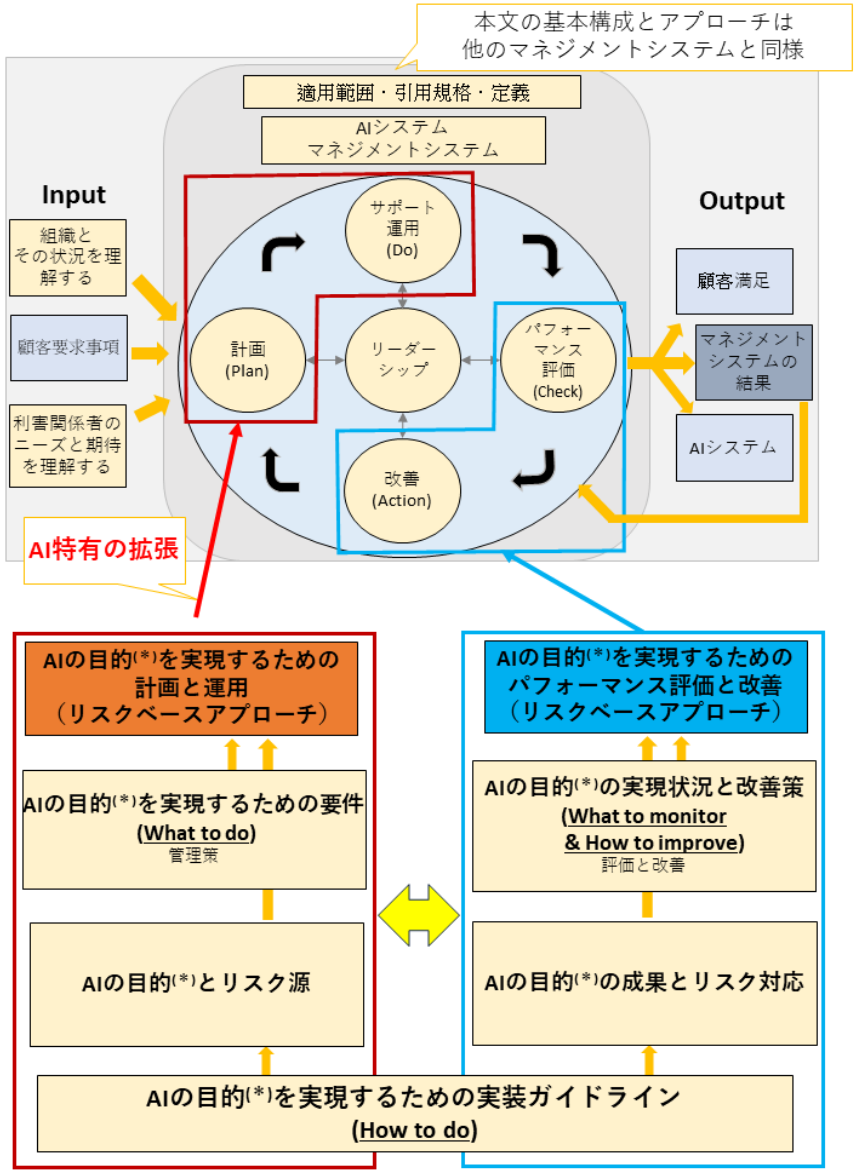
- 近年、AIの開発が活発化しており、一般にも、AIシステムとして日常生活の様々な場面で使用されるなど普及が急速に進みつつあります。その普及に当たっては、安全・安心なAIシステムとして適切に開発・提供・使用することが必要であることから、よりどころとなるマネジメントシステムのニーズが高まっていました。
- そのため、ISO/IEC JTC1/SC42において、AIマネジメントシステムの国際標準化の検討・開発が進められ、2023年12月18日に国際規格「AIマネジメントシステム（ISO/IEC 42001）」として発行されました。

■規格の概要

- 本規格は、AIシステムを開発、提供または使用する組織を対象とし、組織がAIシステムを適切に利活用（開発・提供・使用）するために必要なマネジメントシステムを構築する際に遵守すべき要求事項について、リスクベースアプローチによって規定したものです。信頼性や透明性、説明責任を備えたAIシステムの利活用ができるよう、そのリスクを特定し、軽減すると共に、AIの公平性や個人のプライバシーなどへの配慮についても要求しています。さらに、AIシステムに特有な学習データや機械学習について考慮するにあたっても、重要な規格になります。
- マネジメントシステムの構築については、ISO 9001やISO/IEC 27001規格など既存のマネジメントシステム規格と同様のアプローチを採用し、同じ構成で要求事項を規定するなど、本規格の利用者を考慮した規格にもなっています。
  - (METIニュースリリース2023年1月15日)

■期待される効果

- 本規格により、AIシステムの開発・提供・使用をする事業者が国際標準に基づいたAIマネジメントシステムを構築し、これまで以上に安全・安心なAIシステムの開発・提供・使用が行えるようになることが期待されます。また、AIシステム関係者相互の共通理解が図られるようになり、AIシステムの国際取引が促進されることも期待されます。
- 図：AIマネジメントシステムの構成（METIニュースリリース2024年1月15日）
  https://www.meti.go.jp/press/2023/01/20240115001/20240115001.html



＊AIの目的：組織が開発・提供・使用するAIで達成したいこと

3.4 ISO 42001 目次

Introduction

1.Scope

2.Normative references

3.Terms and definitions

4.Context of the organization
- 組織とその背景、関係者ニーズと期待、AIマネジメントシステムの範囲

5.Leadership
- リーダーシップとコミットメント、ポリシー、役割/責任/権限

6.Planning
- リスクと機会に対処するための行動、AIリスク評価と対処、影響評価、目標達成計画、変更管理

7.Support
- リソース、能力、意識、コミュニケーション、文書

8.Operation
- 運用計画と管理、AIリスクアセスメント、リスク評価

9.Performance evaluation
- モニタリング、測定、分析、評価、内部監査、マネジメントレビュー

10,Improvement
- 継続的改善、不適合の是正措置

Anx.A (normative) Reference control objectives and control - 参照管理目標及び管理

Anx.B (normative) Implementation guidance for AI controls - AI 管理の実施指針

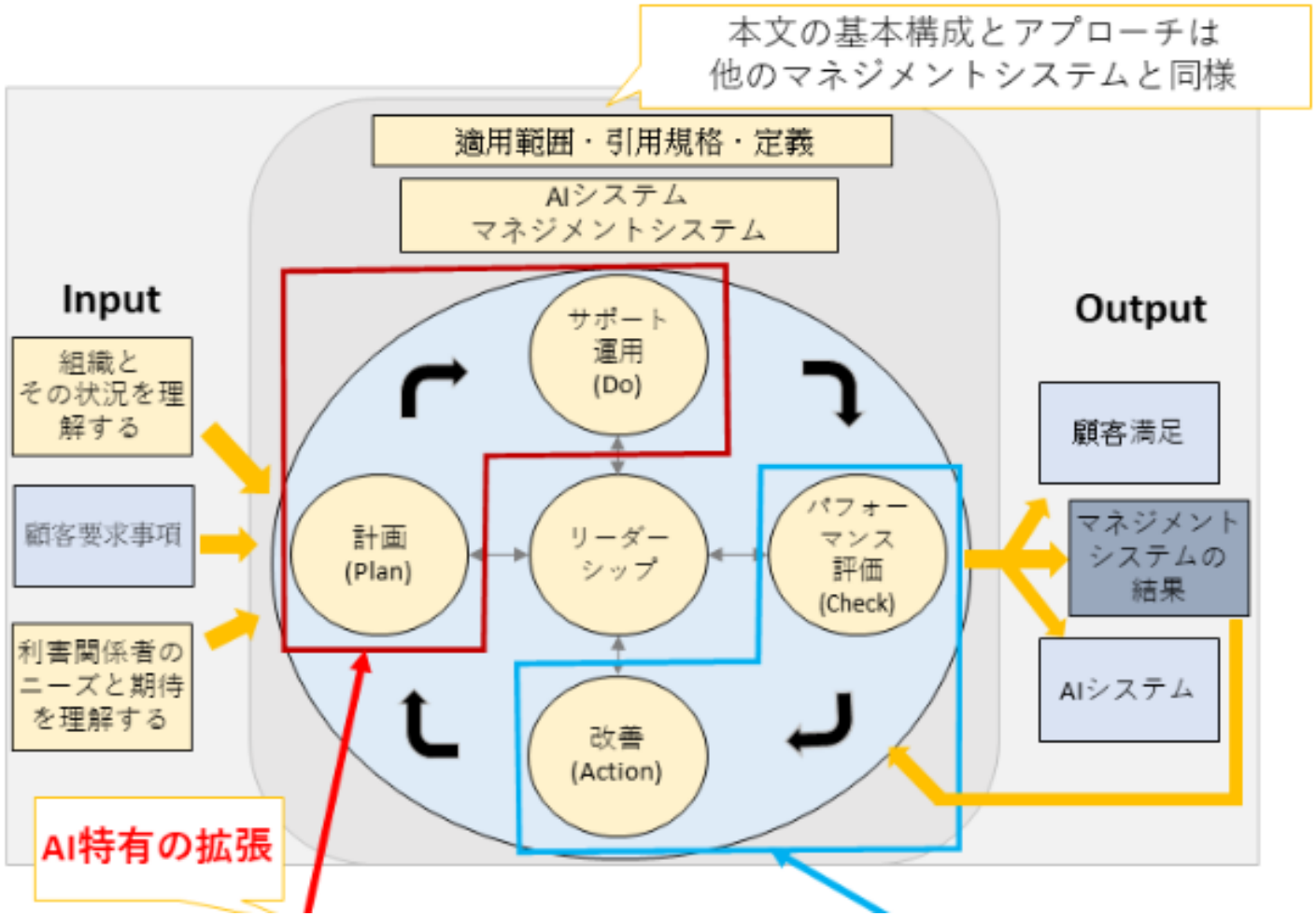
Anx.C (informative) Potential AI-related organizational objectives and risk sources - AI関連の組織目的及びリスク源

Anx.D (informative) Use of the AI management system across domains or sectors - ドメイン又はセクターを超えたAIマネジメントシステムの使用

■特徴

- ISO 9001等のマネジメントシステムをAI固有の拡張
  - マネジメント範囲の定義
  - マネジメント方針の策定
  - マネジメント計画(Plan)
  - プロセス管理，リソース管理，情報管理(Do)
  - パフォーマンス評価(Check)
  - 管理（計画）の改善(Action)

- リスクベースアプローチ
  →AIの目的を実現するための実装ガイドライン



3.6 Annex A.管理目標と管理の参考

■A.1 一般

- 表 A.1 に詳述する管理は，AI システムの設計及び運用に関連するリスクに対処するための組織目標を達成するための参考となる。AIシステムの設計及び運用に関連する組織目標を満たし，リスクに対処するための参考となる。
- 表A.1に記載されているすべての管理目標及び管理策を使用しなければならないわけではない。また、組織は独自の管理を設計し、実施することができる（6.1.3参照）。
- 附属書Bは、表A.1に列挙されたすべての管理策の実施指針を提供している。

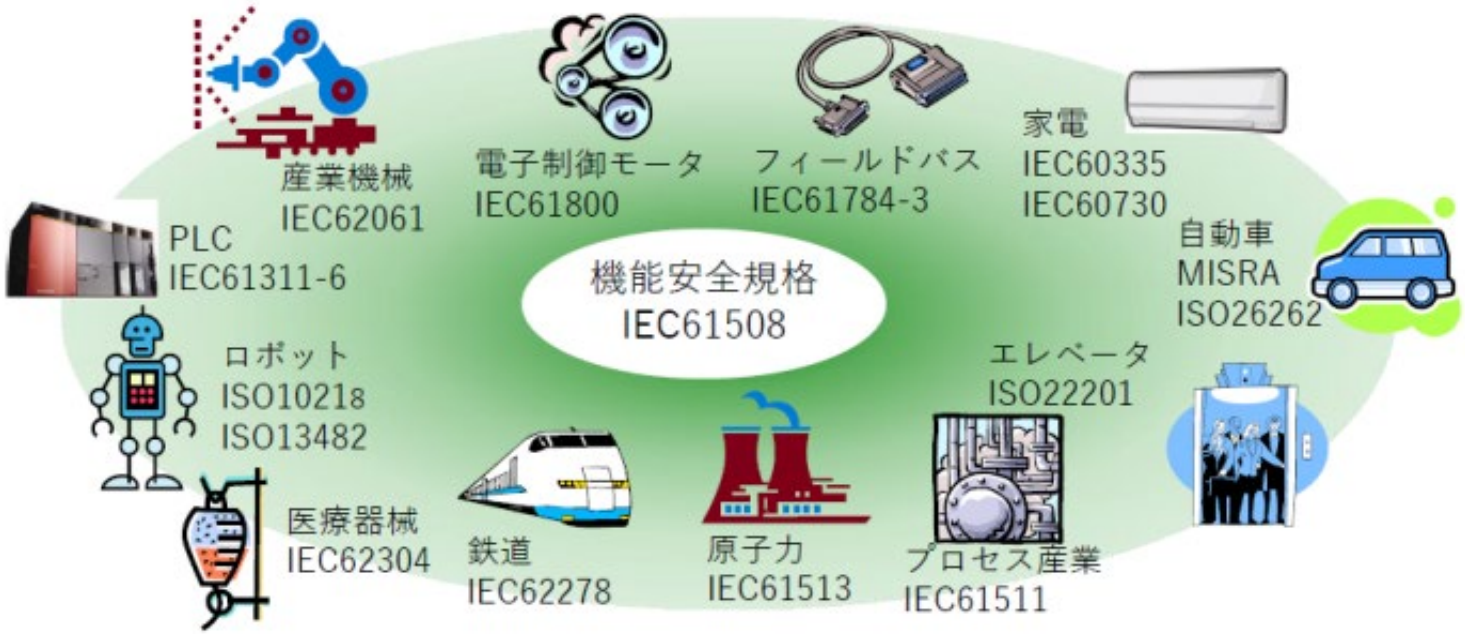
表A.1 - 管理目的及び管理策

A.2 AIに関する方針 目的：ビジネス要求事項に従って、AI システムに対するマネジメントの方向性と支援を提供する。		
	項目	管理
A.2.2	AI方針	組織は、AIシステムの開発又は使用に関する方針を文書化しなければならない。
A.2.3	他の組織の方針との整合性	組織は、AIシステムに関して、他の方針が組織の目的の影響を受ける可能性がある、又は適用される可能性がある箇所を決定しなければならない。
A.2.4	AI方針の見直し	AI方針は、その継続的な適合性を確実にするために、計画された間隔で、又は必要に応じて、見直しを行わなければならない、 適切性及び有効性を確保する

4 IEC 61508機能安全

■IEC 61508 電気電子プログラマブル装置の機能安全

- 安全システムの安全性要求を初めて規定した規格
- 安全制御系：センサ等の入力→処理部が危険状態を判断→安全アクチュエータ出力
- プロセッサやソフトウェア等のIT技術の導入により、高度で複雑な安全制御系を実現
- 多くの分野規格が開発されて多様な製品が実用化
  - 自動消火装置，自動ブレーキ，緊急停止，運転支援装置etc

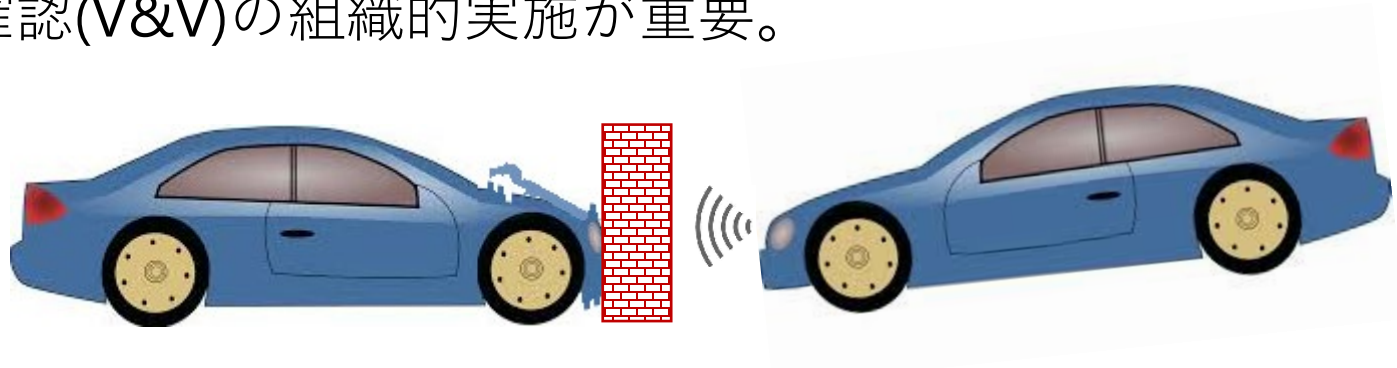


多様な分野における機能安全規格

■安全機能(Safety function)をプログラマブル機器により実現

■機能安全の特徴

- リスクアセスメントに基づいてリスクをSIL1～4に特定。
- リスク低減する安全機能（ハードウェア、ソフトウェア等）の実現にSILごとの要求
  - 高リスク(SIL4)への安全機能は、確実高信頼な手法を用いて実現することが求められる。
  - 例）ハードウェア：故障で安全機能が働かなくなる危険側故障率の目標や、多重化構成や診断率の要求
  - 例）ソフトウェア：バグや設計不具合が発生しないよう設計手法の選択、仕様と試験のトレーサビリティ確保や仕様変更管理などの要求
- 特に、安全に関する瑕疵がないことを担保するための、動作のホワイトボックス化、検証・妥当性確認(V&V)の組織的実施が重要。



(著者オリジナル)

5 ISO/IEC TR 5469

Functional Safety and AI systems

5.1 経緯

■IEC 61508　機能安全規格Ed.3改訂

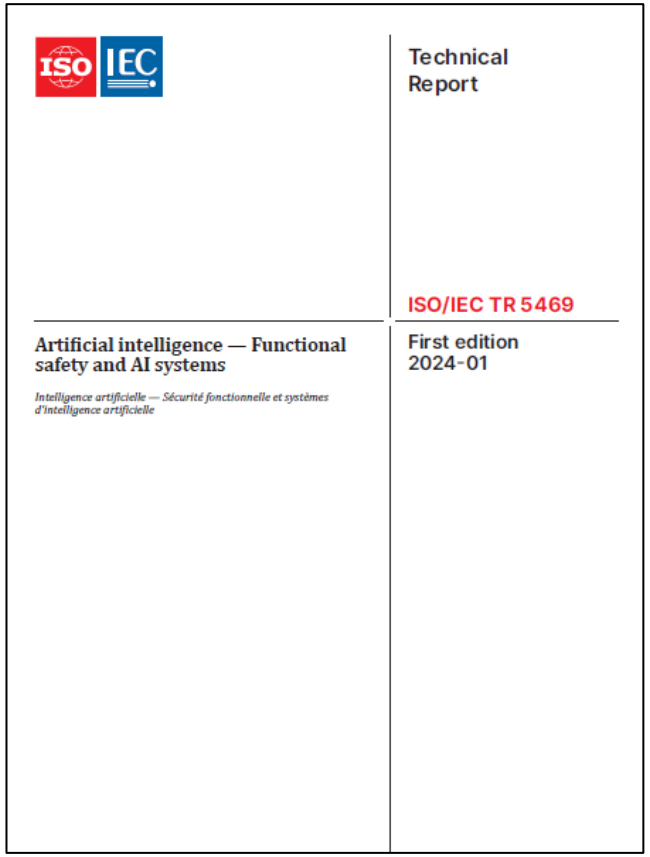
- 電子/電気/プログラマブルシステムによる安全機能の実現のための要求
- 自動運転や協働ロボットのようなAI応用の安全性をどうする？
  - Ed.2では、AIは故障診断を除いて"推奨しない"
- IEC/SC65A/MT61508とJTC1/SC42/WG3のJoint提案（2019 Tokyo)

■ISO/IEC TR 5469 機能安全とAI

- SC42/WG3/FSafetyと、Liason TF(MT61508+WG3有志)、
  コンビナ：Audrey Canning、エディタ：江川(産総研/慶応大)
- 2022年7月：DTR作成と国際コメント募集
- 2023年6月：最終ドラフト(DTR)作成→SC42/WG3にて審議
- 2024年1月10日：ISO/IEC TR 5469:202発行

■EU AI関連法との関係

- AIの安全性に関する唯一の標準→整合規格化を検討



5.2 TR 5469 - Introduction

- 機械学習（ML）などのAI技術で実現される機能の場合、それらが特定の方法で動作する理由を説明し、その性能を保証することは難しい。AI技術が、特に安全関連システムを実現するために使用される場合は、特別注意を払う。
- AI技術は、安全性に影響を与える可能性のあるアプリケーションを実現し始めている。
- AI技術を使用してモデルを継続的に改善する場合、安全性の整合性を実証するための検証および妥当性確認のアクティビティが損なわれる可能性がある。
- この文書の目的は、安全関連システムの開発者が、AI技術の特性、機能安全リスク要因、利用可能な機能安全方法、および潜在的な制約についての認識を促進することにより、安全機能の一部としてAI技術を適切に適用できるようにすることである。このドキュメントは、AIシステムの機能安全に関連する課題とソリューションの概念に関する情報も提供する。
- このドキュメントの目的は、要件を定義することではない。

1. Scope　範囲

●以下に関連するプロパティ、関連するリスク要因、利用可能な方法およびプロセスを説明する。

- 機能を実現するための安全関連機能内でのAIの使用。
- AI制御機器の安全性を確保するための、AI以外の安全関連機能の使用。
- 安全関連機能を設計および開発するためのAIシステムの使用。

2. Normative references　参照規格

3. Terms and difinition　用語と定義

1 safety
2 functional safety
3 functional safety risk
4 organizational risk
5 harm
6 hazard
7 hazardous event
8 system
9 systematic failure
10 safety-related system
11 safety function
12 equipment under control (EUC)
13 programmable electronic (PE)
14 electrical/electronic/programmable electronic
15 AI technology
16 AI model
17 test oeacle

4. Abbreviations　略語

5.1 General一般

- 機能安全は、人の怪我や健康、環境、製品や機器の損害に関するリスクに焦点を当てている。
  - リスクの定義は分野によって異なるが、ここでは第3章（機能安全リスク）の定義を使用。
  - リスクアセスメントは、危害の原因を特定し、製品またはシステムの意図された使用および合理的に予見可能な誤用に関連するリスクを評価する。リスクの軽減は、リスクが許容できるようになるまでリスクを軽減する。許容可能なリスクとは、現在の最先端技術に基づいて、その分野（アプリケーション）で受け入れられるリスクのレベルである。
- IEC 61508シリーズは、3ステップアプローチがリスク低減の優れた実践としている。
  - Step1：本質的に機能的に安全な設計
  - Step2：ガードと保護装置 ←本書の対象
  - Step3：エンドユーザー向けの情報

5.2 Functional safety　機能安全

- IEC61508-4　安全機能→特定の危険なイベントに関して、制御対象の安全な状態を達成または維持することを目的
  - 安全機能は、人や環境に害を及ぼす可能性のある危険に関連するリスクを制御する。
  - また、安全機能は深刻な経済的影響を与えるリスクを減らすことができる。
- 安全機能のハードウェア故障の影響から、システマティック故障に注目
  - システマティック故障：開発中の不具合やバグ等
- 安全機能へのAI技術の使用
  - AI固有のリスク低減策、システマティック故障に注目。リスクと分類に関する考慮事項。
- 高度な安全コンセプトの有効性の保証→要件の増加
  - ただし、AI技術を含む安全機能については、これらの機能を実装するシステムの体系的な機能が使用目的の環境に十分であるという保証にさらに焦点を当てることは避けられない。

6.1 Problem description 課題

- AI技術とAIシステムの使用は、現在、成熟した機能安全国際規格では扱われていない。

6.2 AI technology in safety-related E/E/PE systems 安全関連E/E/PESでのAI技術の使用

- AI技術のアプリケーションの様々なコンテキストに基づいて、安全関連のE/E/PEシステムにおけるAI技術の適用性の分類スキームを提供する。
  - AI技術クラス
  - AIアプリケーションと使用レベル

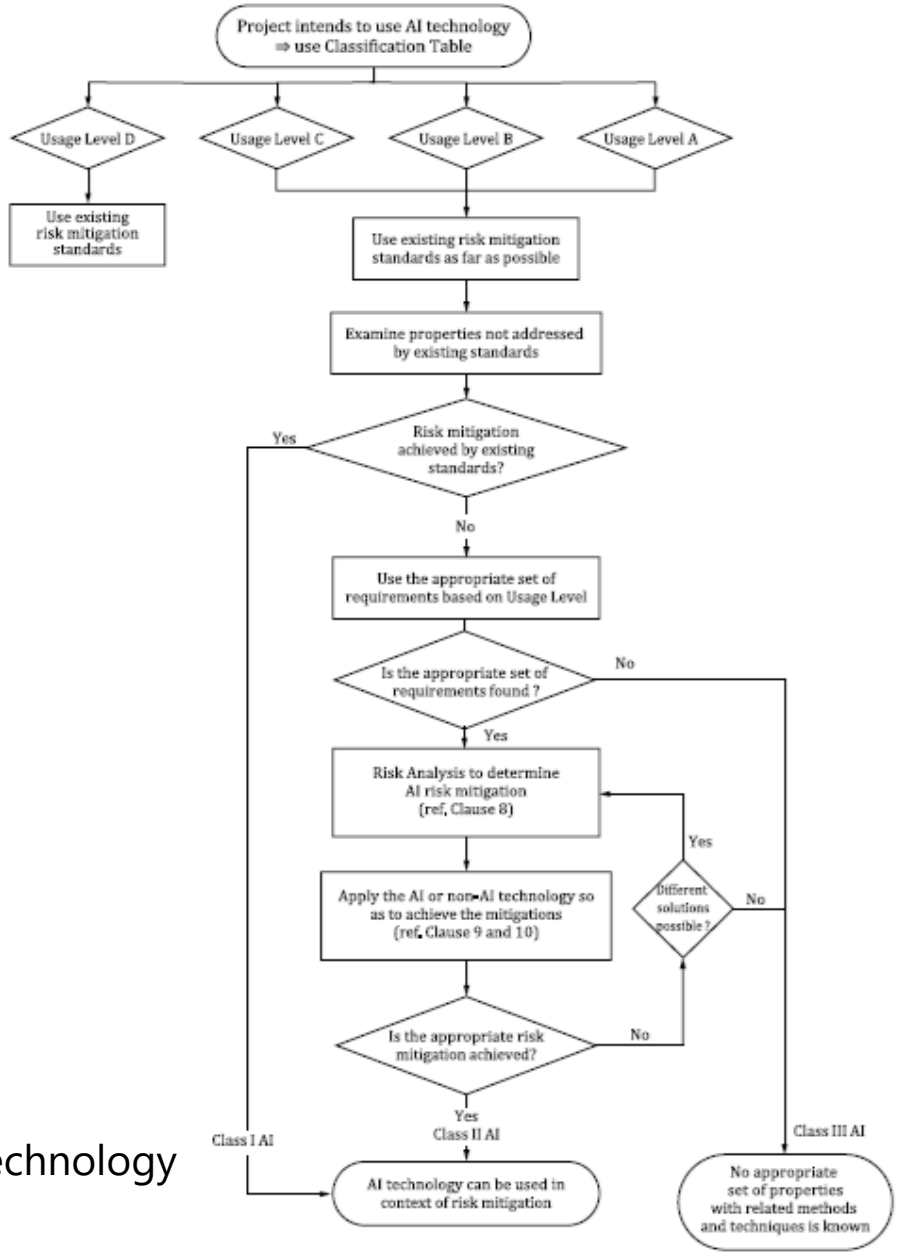


Figure 1 – Flowchart for determining classification of AI technology

図1 AI技術の分類決定のフローチャート

5.7 6.Use of AI technology in safety-related E/E/PE system 安全関連E/E/PESでのAI技術の使用

Table 1 — Example of AI classification table　AI分類表の例

技術クラス=>AIアプリケーションと使用レベル	AI technology Class I	AI technology Class II	AI technology Class III
Usage Level A1（1）	可能な既存の機能安全国際規格のリスク低減の概念の適用	適切な一連の要件（5）	推奨しない
Usage Level A2（1）		適切な一連の要件（5）	
Usage Level B1（1）		適切な一連の要件（5）	
Usage Level B2（1）		適切な一連の要件（5）	
Usage Level C（1,3）		適切な一連の要件（5）	
Usage Level D（2）	AI技術に特定の機能安全要件はないが、既存の機能安全国際規格のリスク低減の概念の適用(4)		

1静的（オフライン）（開発中）教育または学習のみ
2動的な（オンライン）教育または学習が可能
3 AI技術は明らかに追加のリスク削減を提供し、その失敗は許容可能なリスクのレベルにとって重要ではない。
4さらに、他の安全面（機能安全手法では対処されていない）は、AIの使用によって悪影響を受ける可能性がある。
5各使用レベルの適切な要件のセットは、条項8、9、10、および11を考慮して確立できる。例は付録Bに記載されている。

AI技術クラス（Ⅰ～Ⅲ）：識別された一連のプロパティを満たす際のAI技術の実現レベル
AIアプリケーションと使用レベル（A～D）：AI技術の適用、特に、AI技術の使用方法を含む

5.8 7.AI technology elements and the three-stage realization principle AI技術要素と三段階実現原則

Table 2 - AI technology elements
表2-AI技術要素

AI technology element AI技術要素
AI services AIサービス
Machine learning 機械学習 - Model development and useモデルの開発と使用 - Toolsツール - Data for machine learning機械学習のデータ
Engineeringエンジニアリング - Model development and use モデルの開発と使用 - Toolsツール
Cloud and edge computing and big data and data sourcesクラウドおよびエッジコンピューティングとビッグデータおよびデータソース
Resource pool-compute, storage, network ソースプール-コンピューティング、ストレージ、ネットワーク
Resource management-resource provisioning リソース管理-リソースプロビジョニング

Table 3 - Example technology elements involved in model creation and execution for ML
表3-MLのモデルの作成と実行に関連する技術要素の例

技術要素	例（網羅的ではない）
Application graph アプリケーショングラフ	General eXchange Format (GXF) graph in YAML Ain't Markup Language (YAML), recently qualified teacher (rqt) graph in Robot Operating System (ROS)
Machine learning framework 機械学習フレームワーク	TensorFlow, PyTorch, Keras, mxnet, Microsoft Cognitive Toolkit (CNTK), Caffe`, Theano
Machine learning model 機械学習モデル	Open Neural Network Exchange (ONNX), Neural Network Exchange Format (NNEF)
Machine learning graph compiler 機械学習グラフコンパイラ	TensorRT, GLOW, Multi-Level Intermediate Representation (MLIR), nGraph, Tensor Virtual Machine (TVM), PlaidML, Accelerated Linear Algebra (XLA)
Machine code compiler 機械語コンパイラ	Gcc, nvcc, clang/llvm, SYCL, dpc++, OpenCL, openVX
Libraries ライブラリ	CUDA C++, XMMA/SASS kernels, NumPy, SciPy, Pandas, Matplotlib, CUDA Deep Neural Network (cuDNN), SYCL DNN, oneAPI Deep Neural Network (OneDNN), Math Kernel Library (MKL)
Executable machine code 実行可能マシンコード	aarch64, PTX, RISC-V, AMD64, x86/64, PowerPC
Computational hardware	GPU, CPU

7.1 Technology elements for AI model creation and execution AIモデルの作成と実行のための技術要素

- 一部の技術要素は、機能安全の既存の概念で対処できる。
  - 付属書A：機能安全IEC 61508-3の既存の概念をAI技術に適用する方法の例。
  - 付属書B：特定のプロパティ（第8章）を、機能安全の既存の概念を適用できないAI技術に適用する方法の例が含まれる。
- モデルの作成と実行には、さまざまな技術要素が含まれる。

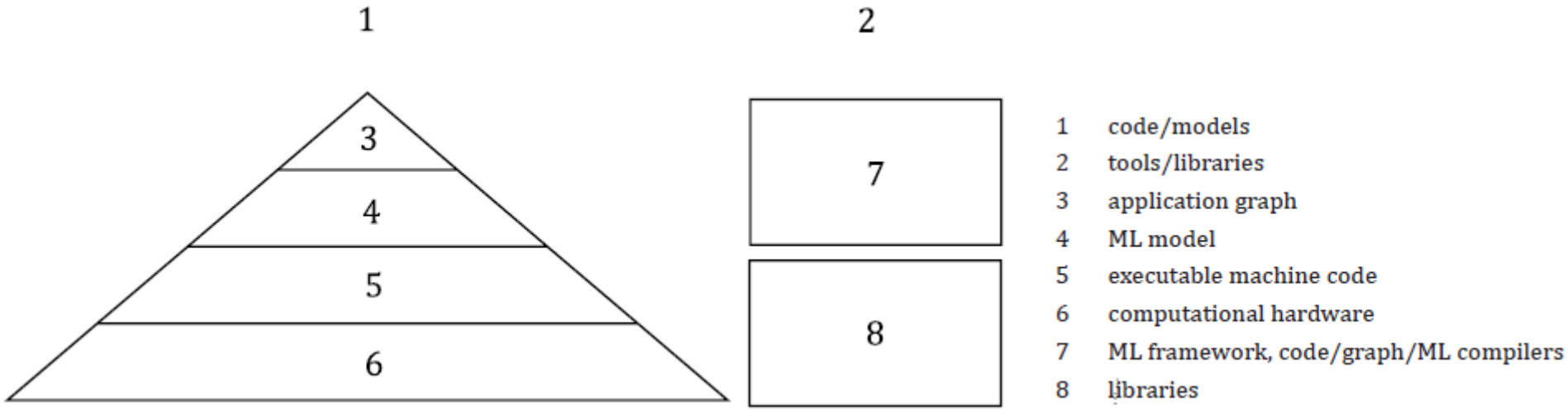


Figure 2 - Hierarchy of technology elements (ML example)

図2-技術要素の階層（MLの例）

7.2 The three-stage realization principle of an AI system AIシステムの3段階の実現原則

- AIシステムは、次の3つで構成される実現原則によって表すことができる（図3）。
  - data acquisition;データ収集
  - knowledge induction from data and human knowledge　データと人知識からの知識の誘導
  - processing and generation of outputs　出力の処理と生成

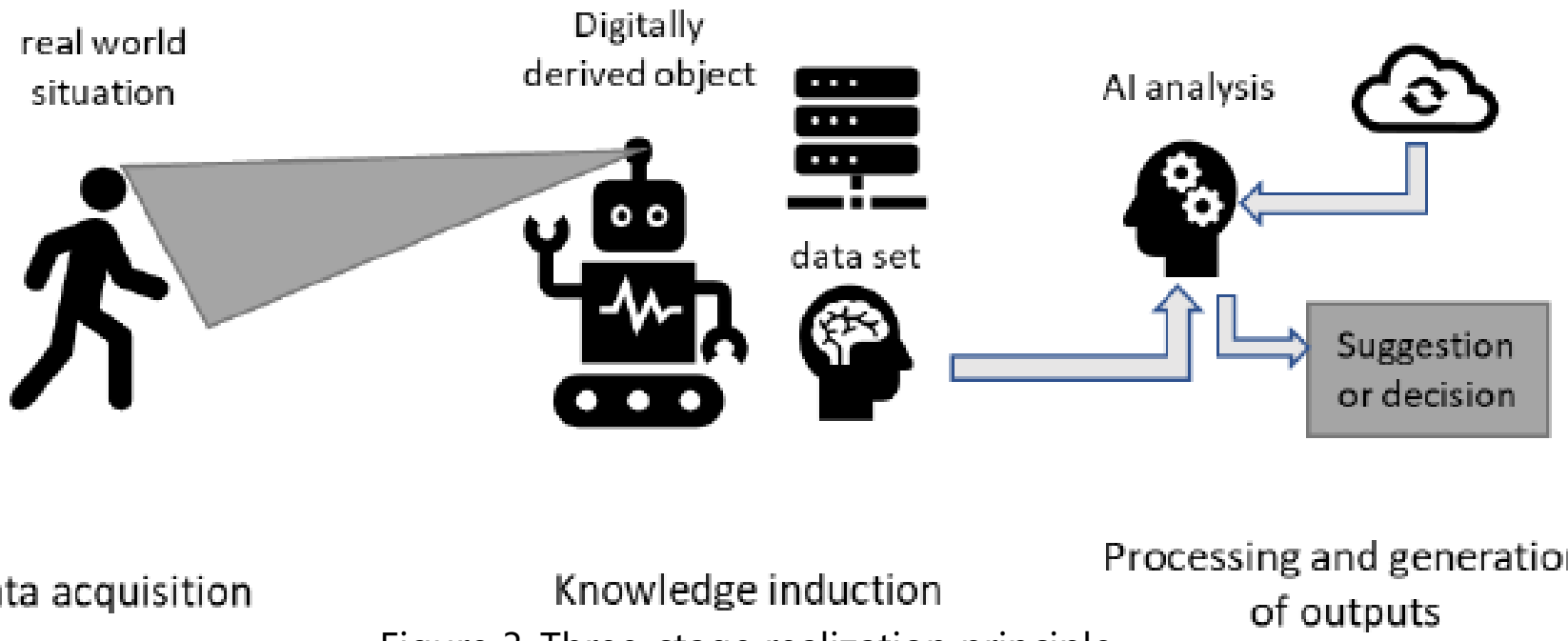


Figure 3-Three-stage realization principle

3段階の実現原則

7.3 Deriving acceptance criteria for the three-stage of the realization principle 実現原則の3段階の受け入れ基準の導出

- 図4のプロセスを定義して、3段階の実現原則に基づいた受け入れ基準を導く。
  - 3つのステージのそれぞれについて、望ましいプロパティを定義。
  - プロパティはトピックに関連し、最終的にトピックに対処する詳細な方法と手法を示す。
  - 合格基準は、一連の詳細な方法と手法から特定される。

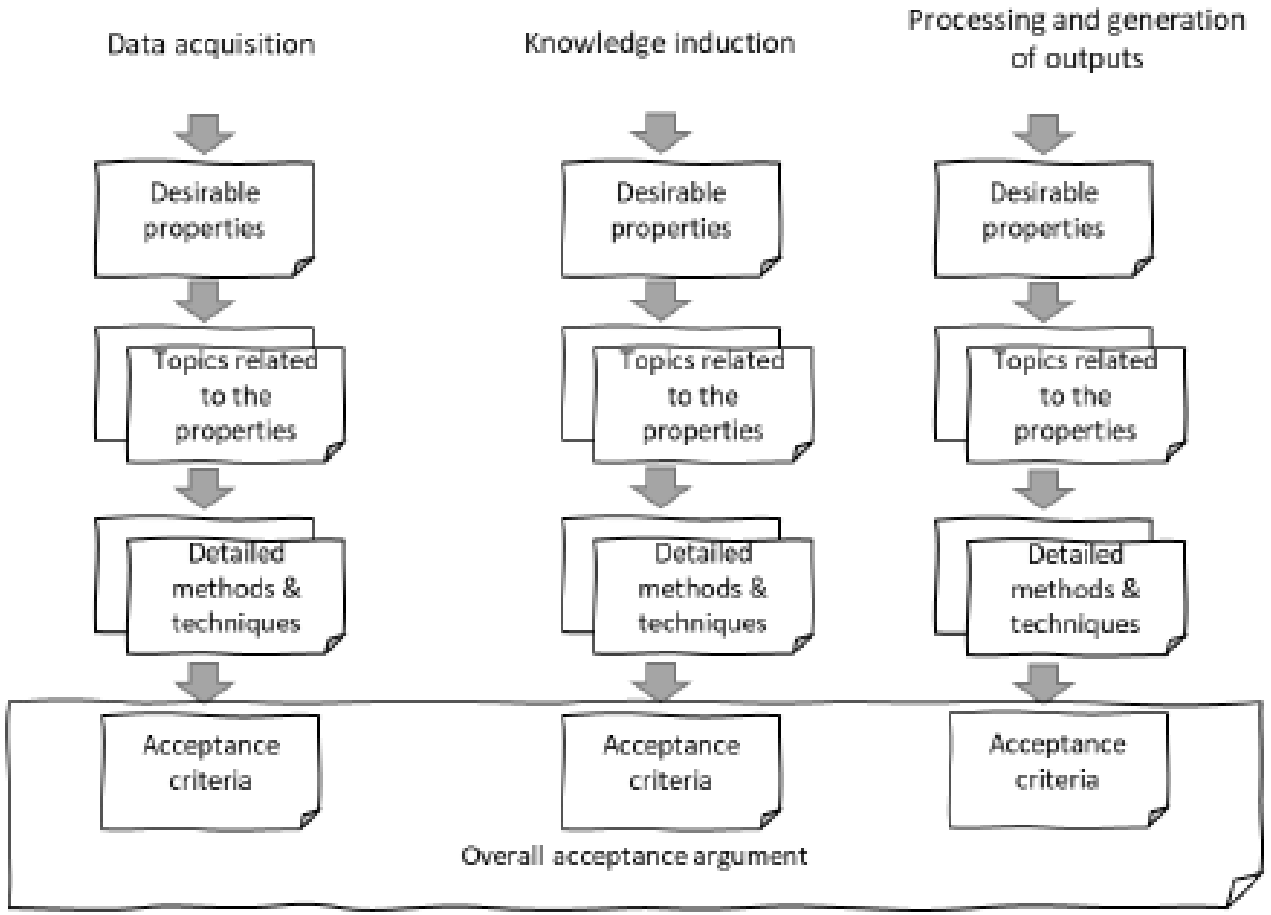


Figure 4 - Processes in each stage　各段階のプロセス

8.Properties and related risk factors of AI systems
AIシステムの特性と関連するリスク要因

8.1 Overview 全体

8.1.1 General　一般

- システムを特徴付けるプロパティと、リスク要因に関するガイダンスを提供する。

8.1.2 Algorithms and models　アルゴリズムとモデル

- AIの機能は、アルゴリズムとモデルの組み合わせによって実現される。
  - アルゴリズム：モデルと入力から情報を推測する。アルゴリズム自体は機能安全の役割を果たさないが、その正確さは機能安全を達成するために非常に重要である。
  - モデル：アプリケーションの目的を達成する情報を表す。機能安全を含むシステムの目的に関連する知識が含まれていることがある。

8.2 Level of automation and control 自動化と制御のレベル

- 自動化レベルは、AIシステムが人間の監視と制御から独立して機能する程度を表す。
  - 高度に自動化されたシステムは、信頼性と安全性の面でリスクをもたらす可能性がある。
- 「スーパーバイザー」は、システムの自動決定を検証または承認するのに役立つ。
- AIシステムの適応性も考慮する。

8.Properties and related risk factors of AI systems
AIシステムの特性と関連するリスク要因

Table 4 - Relationship among autonomy, heteronomy and automation (derived from ISO/IEC 22989, Table 1).

表4-自律性、他律性、自動化の関係

		Level of automation 自動化レベル	Comments
Automated system 自動化システム	Autonomous 自律型	Autonomy 自律性	The system is capable of modifying its operating domain or its goals without external intervention, control or oversight. システムは、外部の介入、制御、または監視なしに、その運用分野またはその目標を変更できる。
	Heteronomous 他律的	Full automation 完全自動化	The system is capable of performing its entire mission without external intervention. システムは、外部の介入なしにその任務全体を実行できる。
		High automation 高度な自動化	The system performs parts of its mission without external intervention. システムは、外部の介入なしにその任務の一部を実行する。
		Conditional automation 条件付き自動化	Sustained and specific performance by a system, with an external agent being ready to take over when necessary. システムによる持続的かつ特定の性能。必要に応じて外部エージェントが引き継ぐ準備ができている。
		Partial automation 部分的な自動化	Some sub-functions of the system are fully automated while the system remains under the control of an external agent. 一部サブ機能は完全に自動化されているが、システムは外部エージェントの制御下にある。
		Assistance 支援	The system assists an operator. システムはオペレーターを支援する。
		No automation 自動化なし	The operator fully controls the system. オペレーターはシステムを完全に制御する。

8.3 Degree of transparency and explainability 透明性と説明性の程度

- 説明可能性をAIシステムのプロパティとして定義し、AIシステムの結果に影響を与える重要な要素を人間が理解できる方法で表現する[ISO/IEC 22989]。
- 透明性は、内部プロセスに関する適切な情報を提供するシステムのプロパティとして定義される。
- 高レベルの説明可能性は、システムの予測できない動作から保護するが、現在の説明可能性技術の制限により、意思決定の品質に関して全体的性能が低下する場合がある。
- AIシステムの一般的な評価に透明性と説明性を含める場合の考慮事項；
  - システムに関する十分な情報が利用可能かどうか。
  - 理解できるか、少なくとも理解できる情報を（間接的に）目的の受信者に配信するか。
  - 正確で、完全で、再現性のある結果を一貫して提供するかどうか。
- 完全に「説明可能なAI」がすぐに達成できない場合でも、リスクに関してモデルの解釈可能性の系統的かつ正式に文書化された評価を採用できる。

8.4 Issues related to environments 環境に関する問題

8.4.1 Complexity of the environment and vague specifications環境の複雑さと仕様のあいまいさ

- AIシステムの開発ライフサイクルは、漠然とした仕様または漠然とした目標から始める。
- 仕様があいまいなため、機能安全関連の特性を保証する際に困難が生じる可能性がある。

8.4.2 Issues related to environmental changes　環境変化に関連する問題

8.4.2.1 Data drift　データドリフト

8.4.2.2 Concept drift コンセプトドリフト

- AI技術を含むコンポーネントは、リスク分析のコンテキストでデータドリフトの原因を検査でき、適切な場合は適切な対策を計画できる。
- モデル設計は、未知の入力に対してでも安全な出力を提供することが期待される。

8.4.3 Issues related to learning from environment　環境からの学習に関する問題

8.4.3.1 Reward hacking algorithms　報酬ハッキングアルゴリズム

8.4.3.2 Safe exploration　安全な探査

8.Properties and related risk factors of AI systems
AIシステムの特性と関連するリスク要因

8.5 Resilience to adversarial and intentional inputs 敵対的および意図的な入力に対する回復力

8.5.1 Overview 全体

- AIシステムの信頼性を評価するときは、敵対的な例や敵対的な攻撃などの意図的な入力に対する機能安全行動の整合性を判断することが適切である。

8.5.2 General mitigations 一般的な緩和策

- 機能安全問題が検出された場合にシステムを引き継ぐ監視機能の適用
- ランダムな障害と体系的なエラーの両方の観点から慎重に検討する

8.5.3 AI model attacks: adversarial machine learning AIモデル攻撃：敵対的機械学習

- AIシステムのモデルは、他のタイプのシステムには見られない特定の弱点を示す可能性があるため、機能安全のコンテキストで展開する場合は、さらに精査する必要がある。

8.6 AI hardware issues ハードウェアの問題

8.7 Maturity of the technology 技術の成熟度

- 成熟度の低い新しい技術は、未知のリスクや評価が難しいリスクの可能性がある。

9.Verification and validation technique
検証および妥当性確認手法

9.1 Overview　全体

- AIシステムと非AIシステムの検証と妥当性確認の手法の違い、および機能安全に適用される違いから生じる問題を解決または軽減するための考慮事項について説明する。
- この条項は、主にクラスⅡの使用法レベルA1からCに適用することを目的としている。

9.2 Problems related to verification and validation 検証と妥当性確認に関連する問題

9.2.1 Non-existence of an a priori specification 先験的仕様が存在しない

- 事前定義された仕様がないと、検証と妥当性確認、および残存リスクの評価に重大な問題が発生する可能性がある。

9.2.2 Non-separability of particular system behaviour　特定のシステム動作の非分離性

9.2.3 Limitation of test coverage テスト範囲の制限

9.2.4 Non-predictable nature 予測不可能な性質

9.2.5 Drifts and long-term risk mitigations　ドリフトと長期的なリスクの軽減

9.3 Possible solutions 考えられる解決策

9.3.1 General 一般

9.3.1.1 Directions for risk mitigation リスク軽減の方向性

- 検証と妥当性確認の実現に向けた二つの方向性
  - 生成されたモデルを分析して、モデルの予想される動作に関する人間が理解できる知識を抽出。
  - 開発プロセス中にAIシステムがどのように構築されているかを分析することにより、機能安全関連の品質を間接的に説明。

9.3.1.2 AI metrics and safety verification and validation メトリクスと安全性の検証と妥当性確認

- 精度などのメトリックは、多くの場合、AIの検証と妥当性確認フェーズで使用できる。

9.3.2 Relation between data distributions and HARA データ分布とHARAの関係

- データセットが理解されるためには、関連のHARAを論理的に理解することが重要。
- 学習中にシステマティックエラーとランダムエラーの両方が発生する可能性があり、機能安全目標違反を引き起こす可能性がある。

9.3.3 Data preparation and model-level validation and verification データの準備とモデルレベルの妥当性確認および検証

- 学習およびテストデータセットが、HARAに関連して決定された基準を満たしていることを確認すること。このデータ要件は、以下の基準に分類できる（付属書C参照）。
  - すべての機能安全関連シナリオに対応
  - リスク状況のすべての合理的なバリエーションをカバー
  - 学習結果をサポートするのに十分な多様性と量
  - 可能な入力のセットに対して安定したテスト結果

9.3.4 Choice of AI metrics メトリクスの選択

- AI技術を含むシステムの性能とKPIを徹底的に評価できる。

9.3.5 System-level testing システムレベルのテスト

- シミュレータの品質と現実性は、優れた検証と妥当性確認の得るために重要。

9.3.6 Mitigating techniques for data-size limitation データサイズ制限の緩和手法

9.3.7 Notes and additional resources 注意と追加リソース

9.4 Virtual and physical testing 仮想および物理テスト

9.4.1 General 一般

- システムの性能を実証するための効果的かつ客観的な方法は、仮想テストまたはシミュレーション。

9.4.2 Considerations on virtual testing 仮想テストに関する考慮事項

- シミュレーションまたは一般的な仮想テストを導入する場合の検討項目

9.4.3 Considerations on physical testing 物理テストに関する考慮事項

- 実世界の環境または最終的な運用環境でシステムをテストすることで、実際の使用の検証を最も忠実に行うことができる。

9.4.4 Evaluation of vulnerability to hardware random failures ハードウェアのランダム故障に対する脆弱性の評価

- たとえば、ソフトエラーに対するディープニューラルネットワークの脆弱性は低い。
  - 安全な動作につながる障害の割合の評価は、特定のタイプのネットワークに役立つ。

9.5 Monitoring and incident feedback 監視とインシデントフィードバック

- 独自のインシデント統計を使用して、安全性能の継続的な証拠を提供できる。
- 運用設計分野と実際の使用プロファイルを使用して、問題の範囲を定義および制限し、テストの対象範囲のメトリックを作成できる。
- 統計的有意性の考慮事項は、ターゲットの故障率と信頼区間からテストデータセットのサイズとテストカバレッジを導き出す。

9.6 A note on explainable AI 説明可能なAIに関する注記

- 十分に説明可能なAI技術は、うまく実現されれば、現在の機能安全国際標準と同様の方法で機械学習アルゴリズムの機能安全を保証する道を開く。
- 特に、AIシステムの意思決定が実装者の意図と異なる場合に役立つ。

10.Control and mitigation measures
制御および緩和措置

10.1 Overview 全体

- この章では、MLモデルの拡張方法をAIシステムのコンポーネントと見なし、信頼性、可用性、品質の非機能特性を改善する方法について説明する。

10.2 AI subsystem architectural considerations サブシステムのアーキテクチャの考慮事項

10.2.1 Overview　全体

10.2.2 Detection mechanisms for switching 切替の検出メカニズム

- 監視モニターは、AI技術の潜在的に危険な動作を検出できる。

10.2.3 Use of a supervision function with constraints to control the behaviour of a system to within safe limits システム動作を安全な範囲内に制御するための制約付き監視機能の使用

10.2.4 Redundancy, ensemble concepts and diversity 冗長性、合成の概念および多様性

- 冗長性には、構造(空間)、時間(頻度)、機能(情報)または組合せなどの種類がある。

10.2.5 AI system design with statistical evaluation 統計的評価を伴うAIシステム設計

10.3 Increase of the reliability of components containing AI technolog. AI技術を含むコンポーネントの信頼性の向上

10.3.1 Overview of AI component methods AIコンポーネント手法の全体

10.3.2 Use of robust learning ロバストな学習の使用

- ノイズの妨害、デバイスの障害、および悪意のある（敵対的な）入力に対する堅牢性を向上させる方法（例：正則化regularization）。

10.3.3 Optimisation and compression technologies 最適化および圧縮技術

- パラメータと計算の定量化、プルーニング、より単純な代理モデルへの知識の蒸留などの最適化および圧縮技術は、ハードウェア固有の計算加速を実現する。

10.3.4 Attention mechanisms アテンションメカニズム

10.3.5 Protection of the data and parameters データとパラメータの保護

- データとモデルのパラメータは、ハードウェア障害から敵対的な攻撃でのデータポイズニングに至るまで、ランダムで意図的な妨害や損失に対して潜在的に脆弱。

11.1 General 一般

- 機能安全を確保するため、システムのライフサイクル全体を通して安全性を考慮する。

11.2 Relationship between AI lifecycle and functional safety lifecycle AIライフサイクルと機能安全ライフサイクルの関係

- 従来の機能安全ライフサイクルから開始し、安全に影響を与えるAIシステム固有の問題を考慮

11.3 AI phases AIフェーズ

11.4 Documentation and functional safety artefacts 文書と機能安全アーティファクト

- 学習プロセス、データの関連性、学習、検証、テストデータ

11.5 Methodologies　方法論

11.5.1 Introduction　はじめに

11.5.2 Fault model　障害モデル

11.5.3 PFMEA of offline training of AI technology AI技術のオフライン学習のPFMEA

- プロセスFMEAにより、学習プロセス内のバイアスと制限の原因を分析および排除できる。

5.25 Annex 付属書

Annex A Applicability of IEC 61508-3 to AI technology elements
AI技術要素へのIEC61508-3の適用性

Annex B Examples of applying the three-stage realization principle
三段階実現原則の適用例

Annex C Possible process and useful technology for verification and validation
検証と妥当性確認のための可能なプロセスと有用な技術

Annex D Mapping between ISO/IEC 5338 and IEC 61508 series
ISO/IEC5338とIEC61508シリーズ間のマッピング

Bibliography

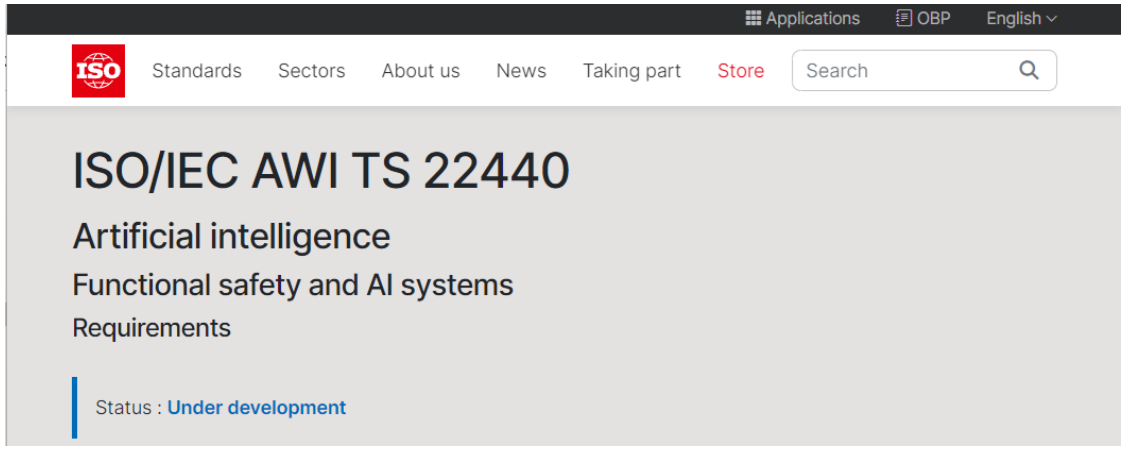
6

ISO/IEC TS 22440

Functional Safety and AI systems -requirements

■ISO/IEC TR 5469 →ISO/IEC TS 22440 Artificial intelligence - Functional Safety and AI systems - Requirements

- Technical ReportからTechnical Specificationへ
- Abstract
  - 以下に関連する用語、特性、リスク要因、プロセス、方法、技術、およびアーキテクチャに関する要件とガイダンスを提供する。
    - 安全関連機能における AI テクノロジーの使用。
    - AI技術を活用したシステムの安全性を確保するために、従来技術に基づく安全関連機能を利用
    - 安全関連機能の設計、開発、検証のための AI テクノロジーの使用。
  - セキュリティの脅威が AI システムの安全性にどのような影響を与える可能性があるかについての一般的な考慮事項が含まれる。
- 欧州AI法「ハイリスク用途」の技術的要求を明確にしたい



6.2 経緯

■TR発行前にNP提案

- IEC/65A/1100/NP　(2023-06-16)
  →成立(2023-8-30: 賛成21，反対1，参加7か国)
- ISO/IEC/JTC1 SC42/JWG4 (AI)とIEC/SC65A/JWG21(FS)の共同作業
  - 2023-10-18 キックオフ会議(ウィーン)
  - 月2回程度のオンライン会議
  - 2026-08 TS発行予定

■現状

- JWGの体制と役割分担の確認
- TS化に向けて各章の見直し担当を募集→ドラフト案
- AIを安全適用したユースケースの募集
- 国内委員会
  - JTC1/SC42/JWG4 - 情報規格調査会/SC42専門委員会/JWG4小委員会
  - IEC/SC65A/JWG21 - 日本電気計測器工業会/TC65諮問委員会/JWG21国内委員会

ISO IEC

65A/1100/NP

NEW WORK ITEM PROPOSAL (NP)

PROPOSER:	DATE OF PROPOSAL:
Secretariat	2023-06-01
DATE OF CIRCULATION:	CLOSING DATE FOR VOTING:
2023-06-16	2023-09-08

IEC SC 65A : SYSTEM ASPECTS

SECRETARIAT:	SECRETARY:
United Kingdom	Ms Stephanie Lavy
NEED FOR IEC COORDINATION:	PROPOSED HORIZONTAL STANDARD: ☐ Other TC/SCs are requested to indicate their interest, if any, in this NP to the TC/SC secretary

FUNCTIONS CONCERNED:
☐ EMC　☐ ENVIRONMENT　☐ QUALITY ASSURANCE　☒ SAFETY

TITLE OF PROPOSAL:
Artificial intelligence — Functional Safety and AI systems — Requirements

6.3 ISO/IEC TS 22440目次

7 まとめ

7 まとめ

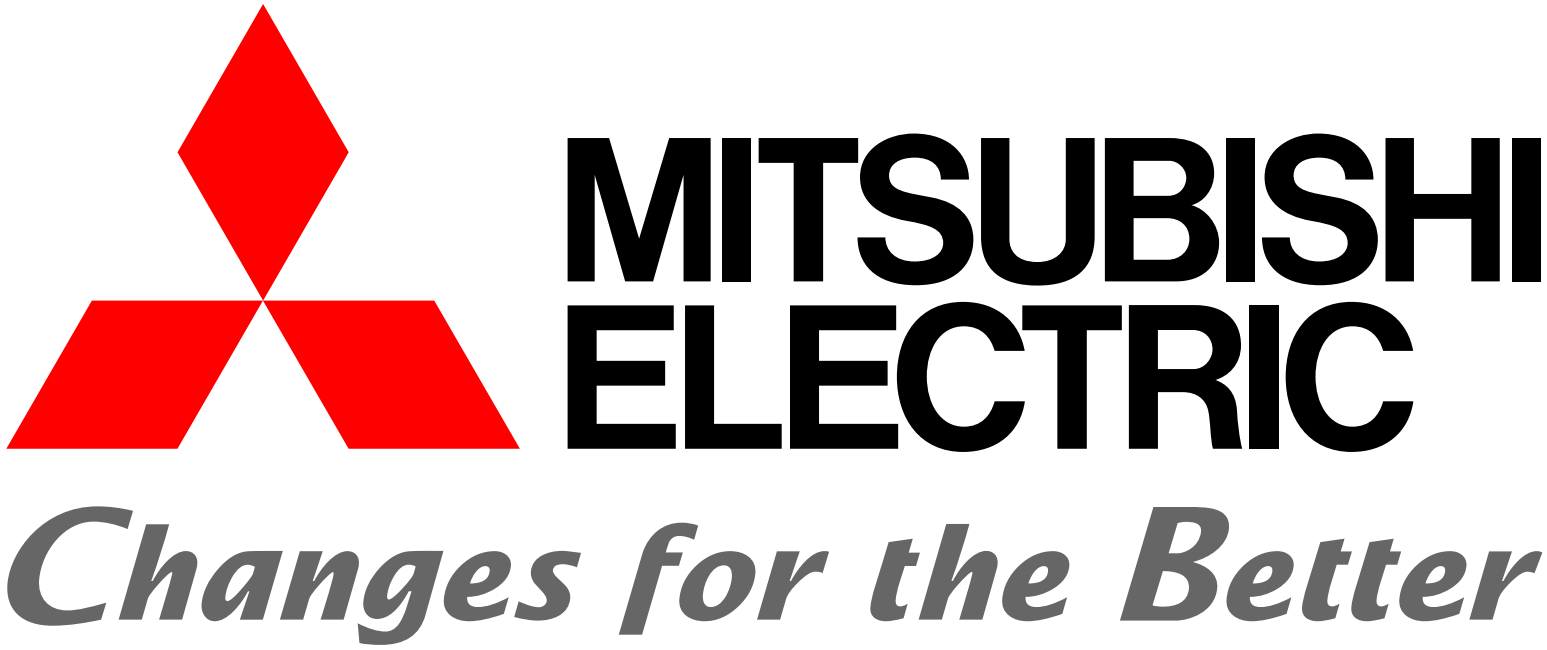
■各国のAI規制

- AI技術を利用した製品・サービスを迅速に市場投入できるように，市場を安心させるための基準・規制
- EU AI法
  - 4段階のリスクレベル（非許容，ハイリスク，限定，最小）に対する事業者/ユーザーへの要求
  - 法令適合，CEマーキング貼付，適合宣言書
  - 2023年12月議会と理事会合意，2027年発効予定

■AI標準化（整合規格に向けて）

- ISO/IEC/JTC1/SC42 AI
  - ISO/IEC 42001:2023 AIマネジメントシステム
  - ISO/IEC TR 5469:2024 機能安全とAIシステム，→ISO/IEC TS 22440 要求事項
  - and more
- 法令が参照する標準に適合必須

■標準化活動に積極的に参加しましょう！


MITSUBISHI
ELECTRIC
Changes for the Better