AIST
Create the Future, Collaborate Together
Digital Architecture Research Center

機械学習品質マネジメントプロジェクトのご紹介

産業技術総合研究所
デジタルアーキテクチャ研究センター
副研究センター長・機械学習品質マネジメントプロジェクトリーダー
大岩 寛

なぜ AI に品質が必要か？

- ソフトウェアによる複雑な制御に人命を預ける時代
  - 航空機・鉄道車両
  - 自動車
  - 消費者機械
  - インフラ（電力等）
  - 医療・ヘルスケア

×

- 機械学習AI等に依存するソフトウェア構築
  - 人間でもルールを記述しきれない複雑な問題への対処
    - 従来型ソフトウェアの構築手法にも限界
  - 逐一変化する実世界への追従

社会からみた AI への恐怖と要求

- 一方で、
  AIの「得体の知れなさ」「社会的影響」への恐怖も顕在化
  - 動作が説明できないブラックボックス的なシステム
  - 「人が作ったものでない」

    ⇒ **規制や合意で制約を掛ける動きが起こる**

  - 社会原則・Principles レベル
  - 法律・社会的ガイドライン レベル
  - 技術ガイドライン・技術標準レベル

社会からみた AI への恐怖と要求

- 2019年頃から: AIに対する社会からの要求の明文化の動き

- **人間中心のAI社会原則**
  (2019. 3　統合イノベーション戦略推進会議)
  - 人間中心の原則
  - 教育・リテラシーの原則
  - プライバシー確保の原則
  - **セキュリティ確保の原則**
  - 公正競争確保の原則
  - **公平性**・説明責任及び透明性の原則
  - イノベーションの原則

- **OECD Principles on AI**
  (2019. 5. 22)
  - 全ての人への普遍的利益
  - **公平性と公正性の確保**
  - 透明性の確保と責任ある開示
  - **堅牢・セキュア・安全性**と
    リスクアセスメント
  - 開発運用者の責任

社会からみた AI への恐怖と要求

- 2020年代: 法律・ガイド層の取り組み

米国

NIST
AIリスクマネジメントフレームワーク作成開始
(2021.7)

米国政府調達要件に入るとサプライチェーンに連なる日本企業にも影響する恐れ

欧州

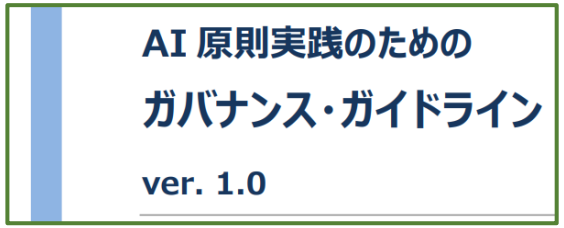
欧州委員会
AI法案を公表
(2021.4.21)

高リスクAI応用を特定
施行されると欧州市場向けビジネスで必須に

日本

経産省
AIガバナンス・ガイドラインを発表
(2021.7.9)

法的拘束力はないが、共通認識形成を通じて、各社の自主的取り組みを促進



ビジネスから見たAIと品質

- 実世界の応用を阻む「品質の問題点」
  1. 顧客に安心して買ってもらえない
     - 一定品質を保証できないので、PoCから先に進めない
  2. 誤動作時に責任を取れない/逃れられない
     - 「無過失」を証明できないので、
       想定外の結果に全責任を負う羽目になりかねない
  3. 売買契約において不利になる
     - 「想定内の瑕疵」と「当初想定を超えた機能拡張」が
       区別できない ⇒ 延々とメンテナンスする羽目になる
     - 安いデータでいい加減に作った手抜きのAIと、
       きちんと慎重に作ったAIが区別できない ⇒ 競争に負ける

機械学習ソフトウェアの特異性

- 機械学習AIはデータから統計的に構築
  - リスク要因を学習させても、
    常に正しく判断するとは限らない
  - 学習結果の構造が判らないので、
    検査をしても網羅性を確保できない
  - 修正をすると、他の所に未知の影響が出る

→ 従来のソフトウェアのための品質管理技術をそのまま適用しても
十分な結果が保証されない ＝ 既存の技術標準だけでは不足している

- **機械学習AIに適合した**
  新たな「品質の作り込み」の枠組みを検討・提案

機械学習品質マネジメントガイドライン

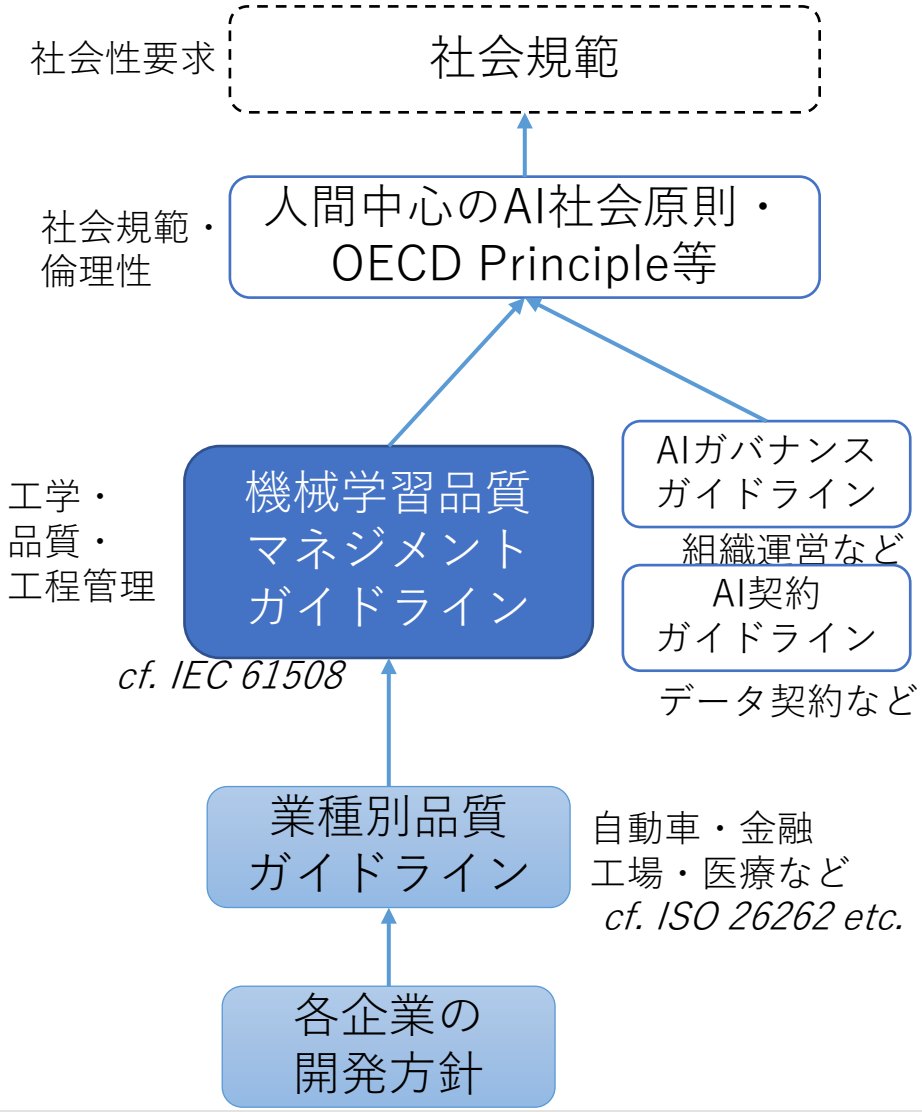
AIST
Create the Future, Collaborate Together
Digital Architecture Research Center

機械学習AIの品質を「作り込み」「確認し」「説明する」ためのガイドライン

- **主な想定読者:**
  - 機械学習を利用して作られる製品やサービスの提供者
  - 実際に製品・サービスをソフトウェアとして実装するシステム開発者
- **2次的な想定読者:**
  - サービス利用者：サービス選択する基準として
  - 第三者評価機関：品質評価・認証の基準として

ガイドラインの位置づけ

- 技術的な側面から
  AIの社会適用を支える
  ガイドライン
  - 社会規範ガイドライン類の下位
    - 「正しさ」の定義はしない
    - 実現する為に**何が必要か**を整理
  - IEC 61508 などに相当
    - 汎用性を持ったガイドライン
    - 業種特有の具体化は有り得る
- 「自社ガイドライン」を
  各企業が作るベース
    - **どう実現するか**には任意性がある



機械学習品質マネジメントガイドライン: 検討体制

- **「機械学習品質マネジメント検討委員会」**
  - 産総研
  - 民間企業10社以上
    - 製品ベンダ・大手ITベンダ・中堅ITベンダー
  - 国立情報学研究所・東京理科大学
  - オブザーバ: IPA・NEDO・METI
- 2018年9月より開催
  - ＋ 週1回程度の詳細検討タスクフォース

取り組みの狙い

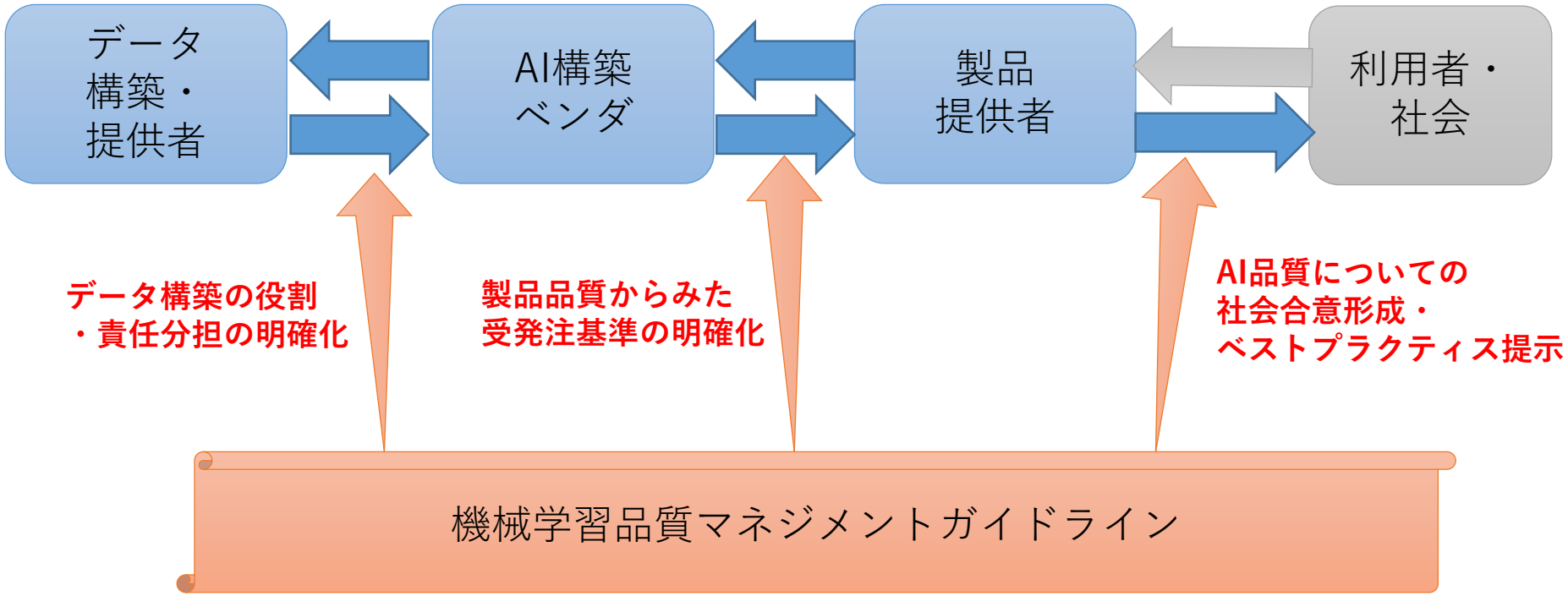
① 社会全体でのAIの受容性向上・安全性向上
- 劣悪なAIの排除による**利用者**の安全性の向上
- 製造物責任の基準明確化による**提供者**のリスク軽減

② AI構築のサプライチェーンの健全性・競争力強化
- AIのサプライチェーン全体での品質管理
- 受発注基準の明確化によるビジネスの障壁除去
- 製品価値のメトリクス提供による日本産AIの競争力の明確化

⇒ **「安心を説明でき、納得して使えるAI」の実現を目指す**

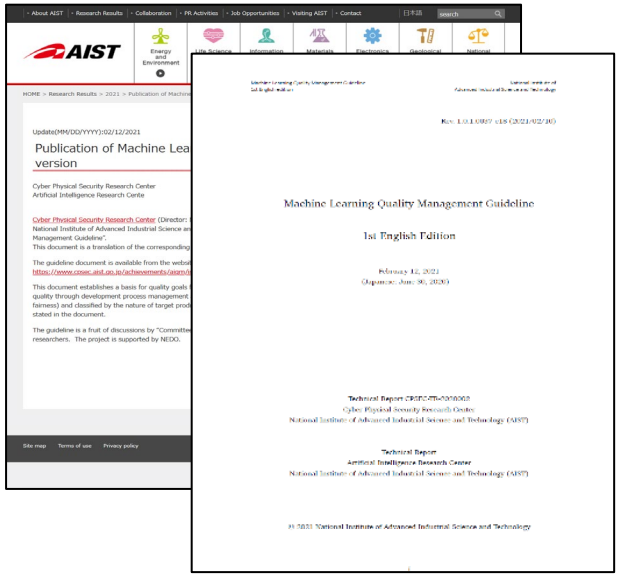
AI品質管理と サプライチェーンマネジメント

- AI 品質を軸にしたサプライチェーン構築支援
  - 品質に関するベストプラクティス・社会合意の整理
  - ステークホルダー間の健全な情報共有・役割分担の形成



ガイドライン：発行実績

- 日本語第1版: 2020年6月公開
- **日本語第3版: 2022年8月公開**
- 英語第2版: 2022年2月公開
- **英語第3版 準備中**



https://www.digiarc.aist.go.jp/publication/aiqm/

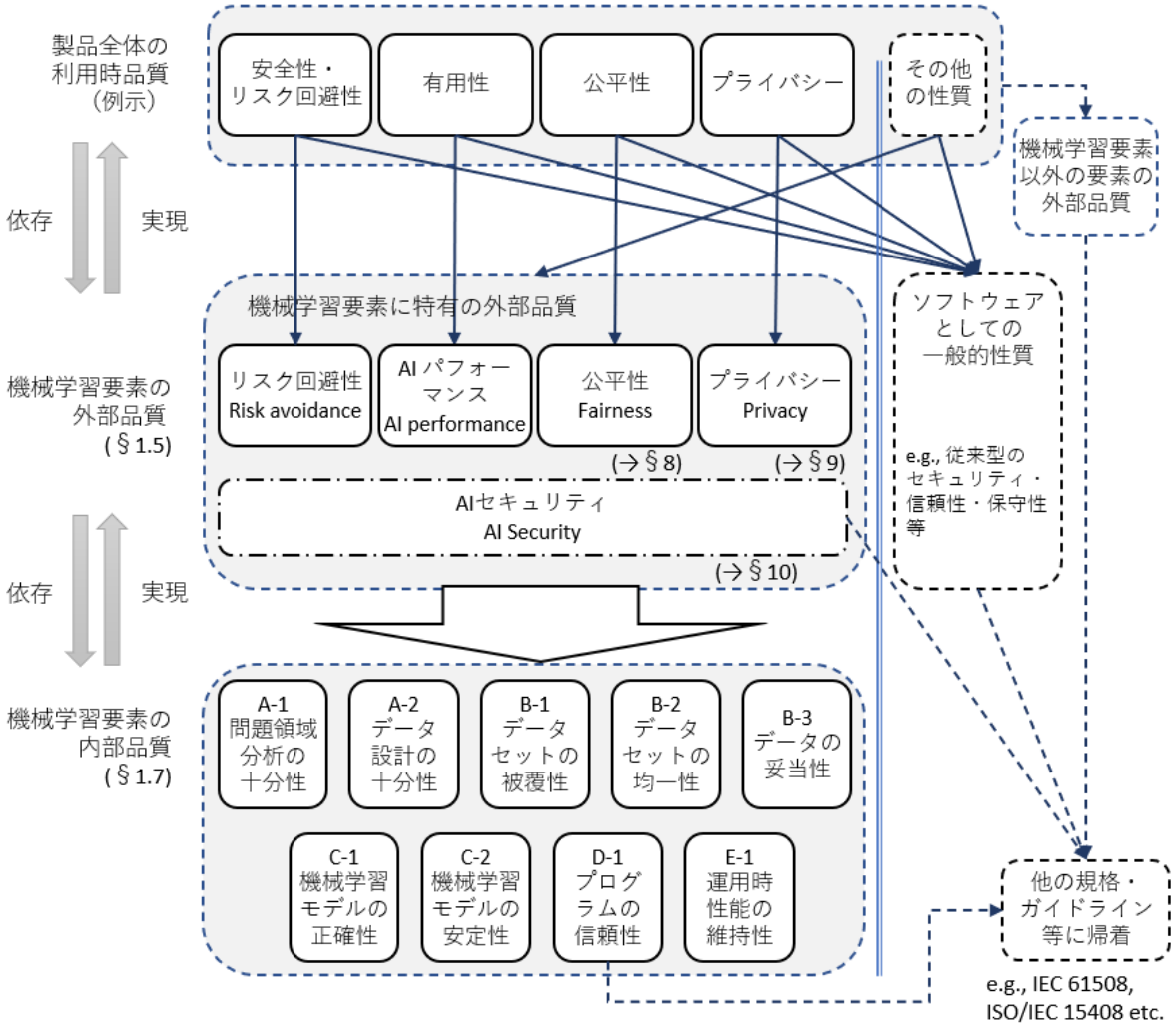
品質管理の対象と考え方

- 着目する「品質」
  - **利用時品質**
    - サービス利用者にとっての品質
      - 安心・公平など
  - **外部品質**
    - システムに「**求められる**」性質

    → **目標を設定**
  - **内部品質**
    - システムが「**持つ**」性質

    → **達成を確認**

  それぞれ**依存・実現**の関係



品質管理の対象と考え方

- 超基本的な流れ
  - 利用者に提供すべき**利用時品質**を考える

  - 製品の設計をして、
    機械学習要素に必要な**外部品質**を考える
  - 外部品質の**品質レベル**を決定する

  - 外部品質レベルに対応した、
    内部品質の要求事項をガイドラインで確認する
  - それぞれの内部品質の達成を確認する

ガイドラインの品質確保の構造

品質目標: 5特性 × レベル
（外部品質特性）

- 機械学習AIが「持つべき**品質**」の目標

① **リスク回避性**（安全性）
  - 7レベルを設定

② **AIパフォーマンス**（トータル性能）

③ **公平性**

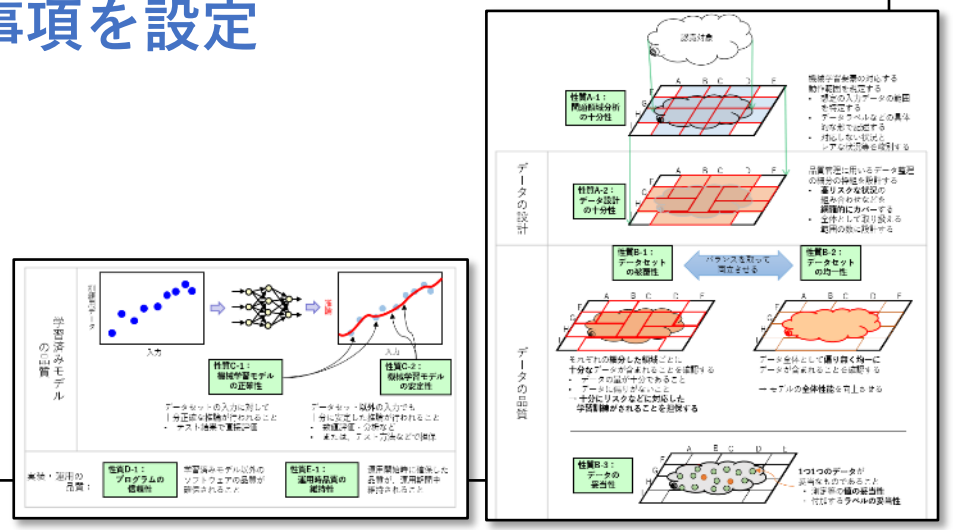
④ **プライバシー**
  - 各3レベルを設定

⑤ **セキュリティ**

品質管理の９ポイント
（内部品質特性）

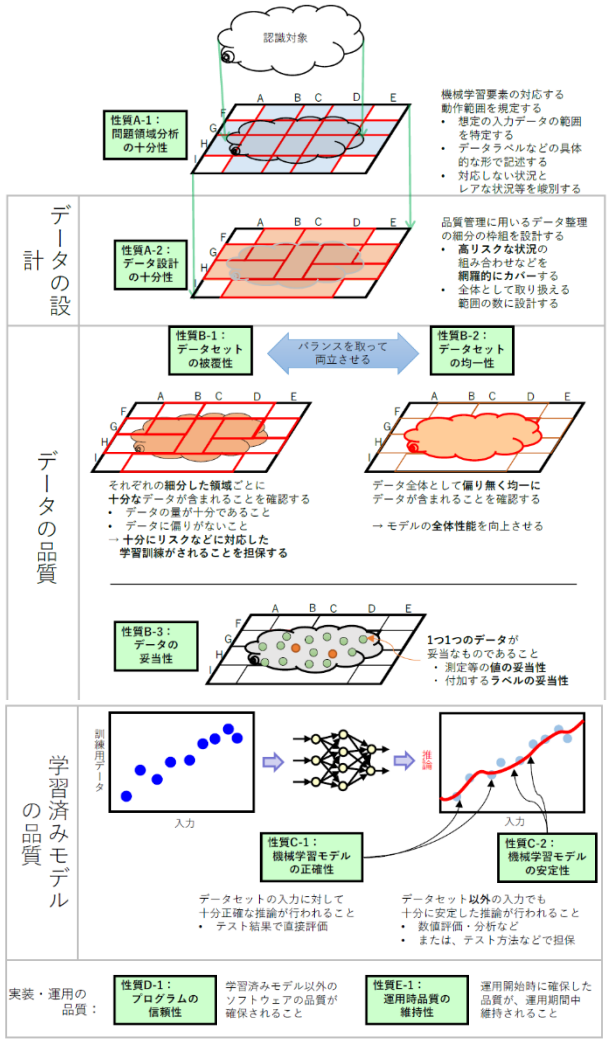
- 品質を向上させるために押さえるべき技術的ポイントを**９項目に整理**

- **左の品質レベルごとに要求事項を設定**



内部品質９項目

A-1. 問題領域分析の十分性
A-2. 問題に対する被覆性
B-1. データセットの網羅性
B-2. データセットの均一性
B-3. データの妥当性
C-1. 機械学習モデルの正確性
C-2. 機械学習モデルの安定性
D-1. プログラムの信頼性
E-1. 運用時品質の維持性



内部品質９項目

基本的な考え方

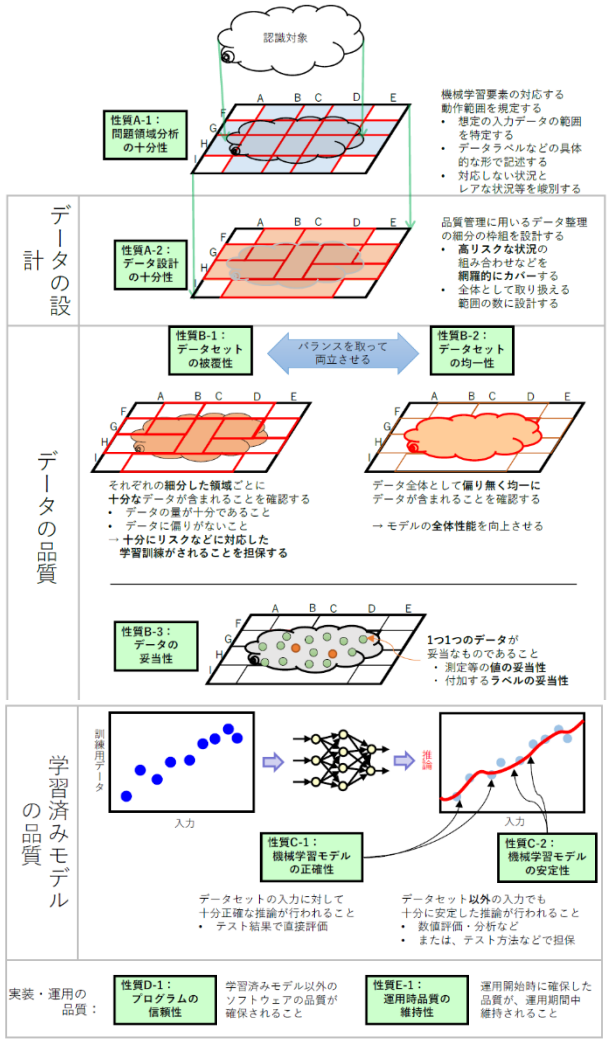
A）問題の分析に基づく
あるべき**データセットの設計**

B）設計に合致する
良いデータセットの確保

C）良いデータセットから得られる
良い機械学習モデル

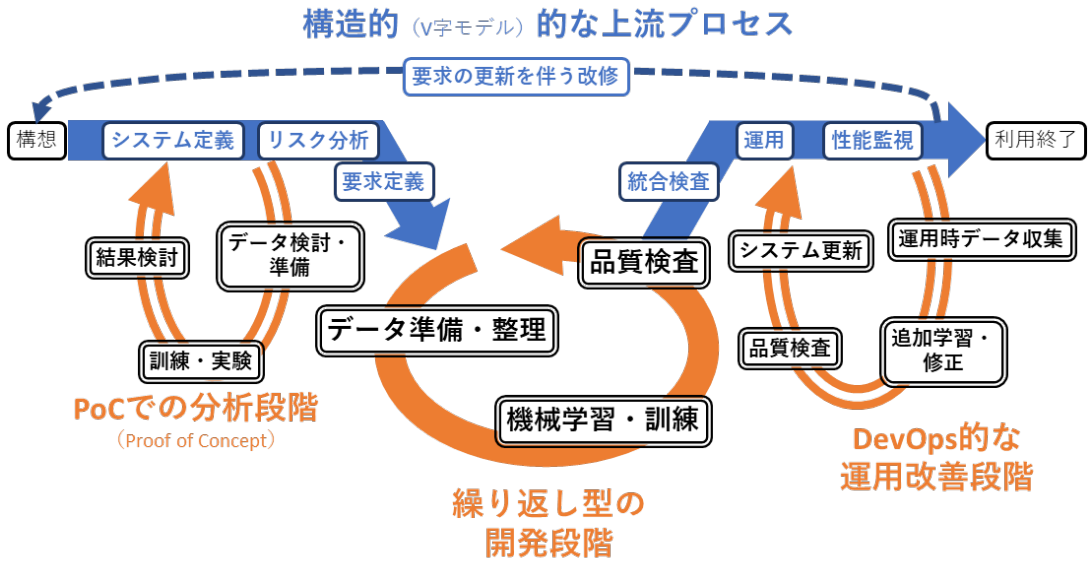
D）信頼できる**ソフトウェア**

E）品質を維持する**運用**



システムライフサイクルプロセス

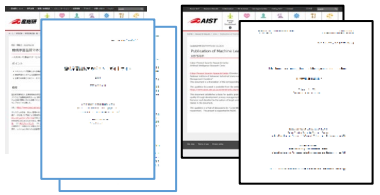
- 品質マネジメントの全体プロセスモデル
  - 企画段階から運用・利用終了までの
    総合的な品質マネジメントを想定して整理
  - AI 特有のPoCプロセスや、繰り返し型の開発工程と、
    品質管理を整合



AI品質ガイドラインの社会展開

機械学習品質マネジメント ガイドライン

- 2022/8 **日本語3版公開**
- 2022/2 英語版2版



AIシステムの品質評価支援 テストベッド

- 2020/11/18 公開

産業展開

民間企業による利用
- ガイドライン適用サポート事業
- 自社AIへの適用

政府系での利用

- **経産省・消防庁・厚労省**
  プラント分野AI信頼性評価ガイドライン
- OECD・EU 協調などでの日本側材料に

国際標準化

- ISO/IEC TR 最新ドラフトに貢献

国際標準化: TR 5469 (Functional Safety and AI systems)

- 担当: ISO/IEC JTC 1/SC 42/WG 3
  - Editor: 日本
  - IEC側 (IEC 61508-3) チームと (事実上合同で) 検討
- Scope:
  - 機能を実現するための安全性機能内部でのAIの利用
  - AIで制御される機器の安全性担保のための非AI機能の利用
  - AIシステムを用いた安全性関連機能の設計と開発
- AIQMガイドラインの内部品質の整理などをインプット
- 投票中

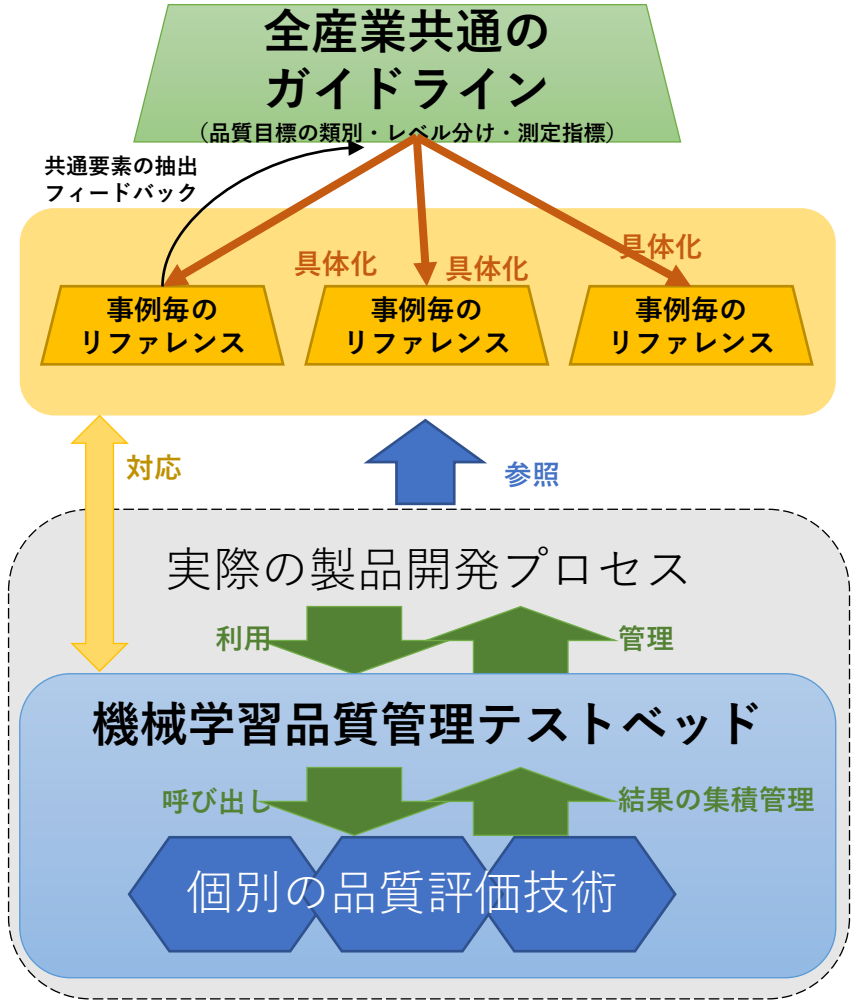
プロジェクトの全体像

①要件の明確化と
エコシステム開発

⇒ 機械学習品質管理ガイドライン
⇒ 産業分野別リファレンスガイド

②実際に品質を作り込む
道具立て
⇒品質管理テストベッド
⇒評価ツール

③ 具体的なAI評価技術の先端開発



品質評価テストベッドQunomon

- **高品質AIを実現するための品質評価環境**
  - 機械学習品質ガイドラインに沿ったテスト管理
  - 豊富な品質テストパッケージと、柔軟な拡張性
  - 品質評価レポートの自動生成

AIテスト技術の流通を促進

発見可能性　入手可能性　相互適合性　再利用性

オープンテストベッド
Qunomon

AIテスト技術を登録 ⇧　⇩ AIテスト技術を利用して
AI品質評価レポートを取得

AIテスト技術開発者　AI品質評価者

2020.11 公開・随時更新中

https://aistairc.github.io/qunomon/

応用別のリファレンスガイド

- ガイドラインを実際に製品に適用するために「こうすればできる」を示す事例集
  - 2022年3月に第1版を公開
- 民間企業の具体的事例を共有知に
  - 民間企業から出向して頂き、産総研で研究

応用別のリファレンスガイド

- 検討内容例
  - データラベルの設計やラベル付けの精度・ツール支援
  - コーナーケースの分析・モデルの安定性などの検証

Road Type	Time of Day	Weather	Pedestrian	Traffic Light	Zebra Crossing	Brightness
General way	dawn/dusk	clear	True	Green	True	Very high
highway	daytime	Cloudy	False	Yellow	False	High
parking lot	night	rainy		Red		Moderate
tunnel	undefined	snowy		None		Low
Under FO		undefined				Very low
undefined						

Table 8: Final problem domain considered in this application

ID	DSA0	DSA1	DSA2	DSA3
73	1.54	0.82	0.97	0.79

ID	DSA0	DSA1	DSA2	DSA3
131	0.52	0.39	1.6	1.15

ID	DSA0	DSA1	DSA2	DSA3
368	4.32	0.98	1.11	0.78

Figure 23: Bike corner case examples detected using Surprise Adequacy

Case	Primary Condition Attribute	Primary Conditional Value	Secondary Conditional Attribute	Secondary Conditional Value	Percentage in the Dataset
0	Weather	Snowy	Road condition	Dry	0
1	Weather	Rainy	Road condition	Dry	0.098
2	Road type	Highway	Signal	Green	1.031
3	Road type	Highway	Signal	Red	0.295
4	Road type	Highway	Signal	Yellow	0.049
5	Road type	Highway	Zebra crossing	Yes	0.344
6	Road type	Highway	Pedestrian	On road	0.098
7	Road type	Highway	Pedestrian	On sidewalk	0

Table 11: Data coverage in unsound cases

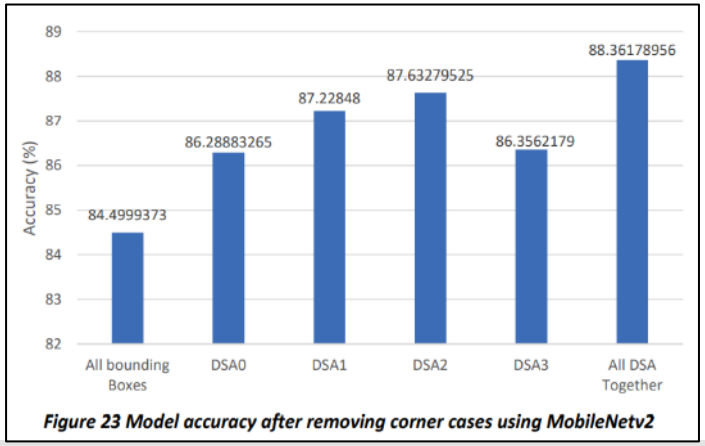
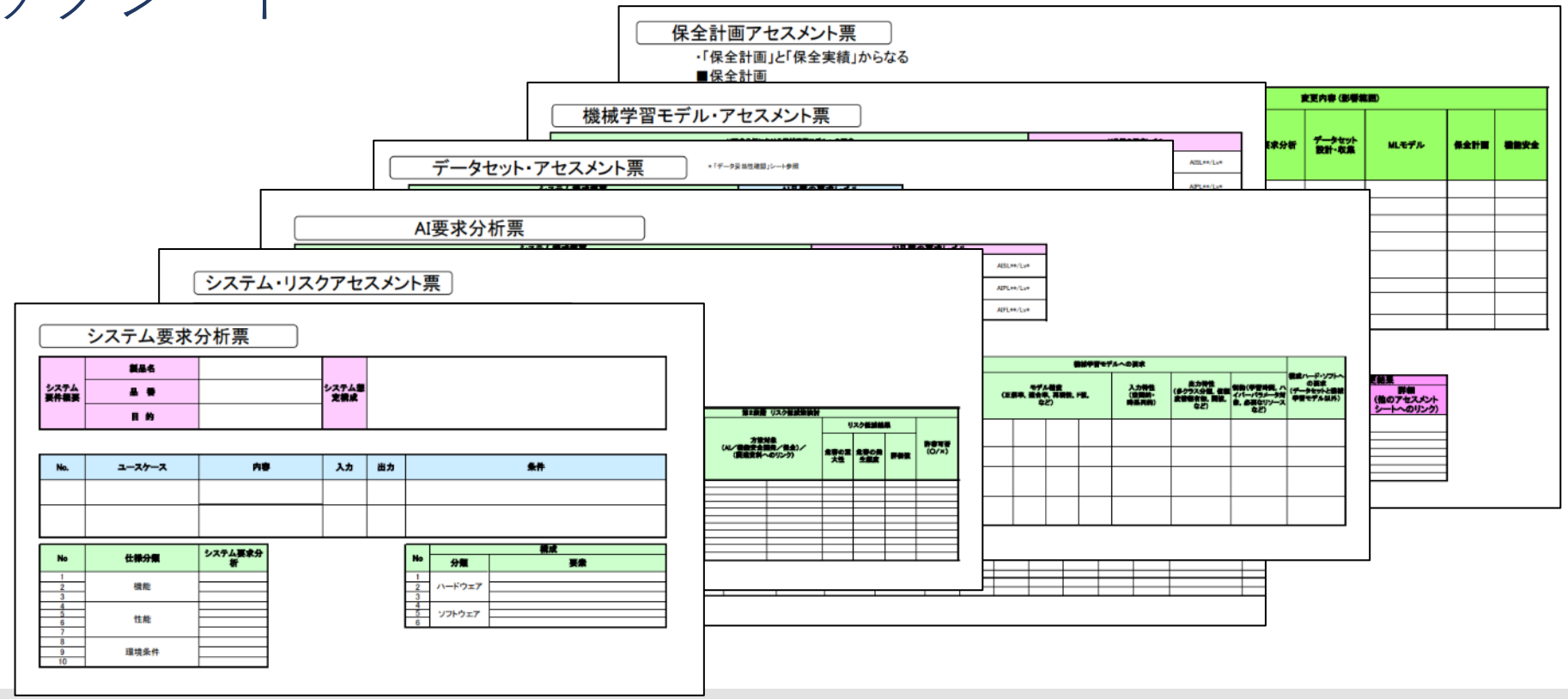


Figure 23 Model accuracy after removing corner cases using MobileNetv2

応用別のリファレンスガイド

- 品質アセスメントシート
  - 製品全体の安全性管理と連結させるための
    チェックシート



まとめ

- 機械学習AIの品質ガイドライン
  - 品質を「作り込み」「**確認し**」「**説明する**」ためのガイドライン

  - サービス提供者とシステム開発者の活用を想定
  - サプライチェーンの確立と利用者・提供者双方の安心感の醸成
  - 品質の基準や認証の基盤としての活用を期待