TOSHIBA

デジタル庁　規制改革推進委員会向け資料

東芝が考えるAI品質保証について

株式会社東芝
代表執行役社長CEO　島田　太郎

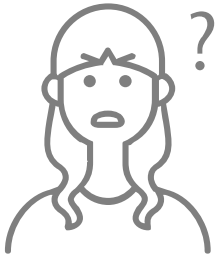
2022年11月16日

01

イントロダクション

AI品質保証は何故難しいか？

品質保証とは



そもそも、品質保証とはなんでしょうか？

品質保証とは、「要求されている性能が、
確かに満たされている」と確認し、示すことです。

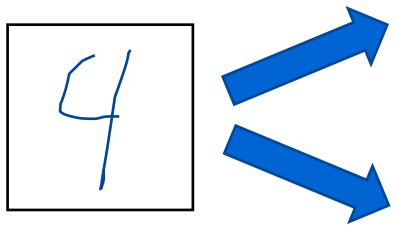
では、なぜAIを使ったシステムは
品質保証が難しいのでしょうか？

AIが持つ不確実性や複雑さが品質保証を
難しくする要因です。詳しく見ていきましょう。

AIを利用する際の品質保証の難しさ

利用者が期待する結果になるとは限らない

予測精度を100%にすることは困難

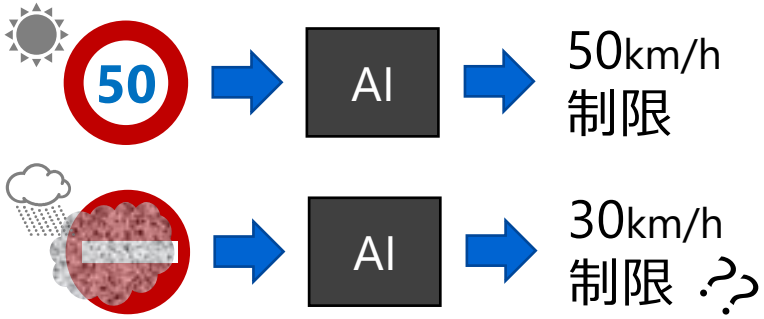


４？

９？

人間でも判断が
難しい問題を
解こうとしている

なぜそのような結果になったか、根拠がわからない



AIは複雑で
ブラックボックス
である

運用開始後の動作を保証することが難しい

動作する環境が変化して正しい予測ができない

時間の経過で
傾向が変わる

学習していないデータが
運用後に出現する

想定していない使われ方をすると結果が予測できない

顔画像認識

？

最も近い人が
推定される？

02

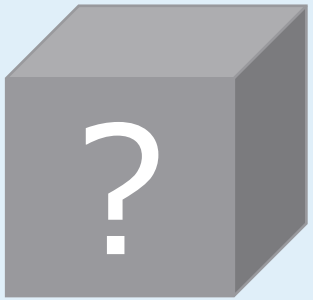
AI品質保証の概要

AI品質保証はAI関連要素全体で行う

AIの課題

AI技術は強力なツールだが従来技術にない課題がある
AI単体は不完全なものであることを前提とすべき

1 ブラックボックス 説明性の確保（透明性）

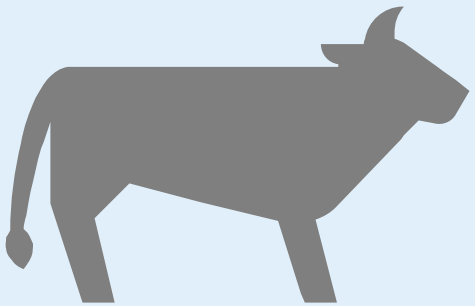


深層学習の判断基準は不明

2 学習データ品質 バイアス解消（公平性）

学生という理由だけで融資を拒否

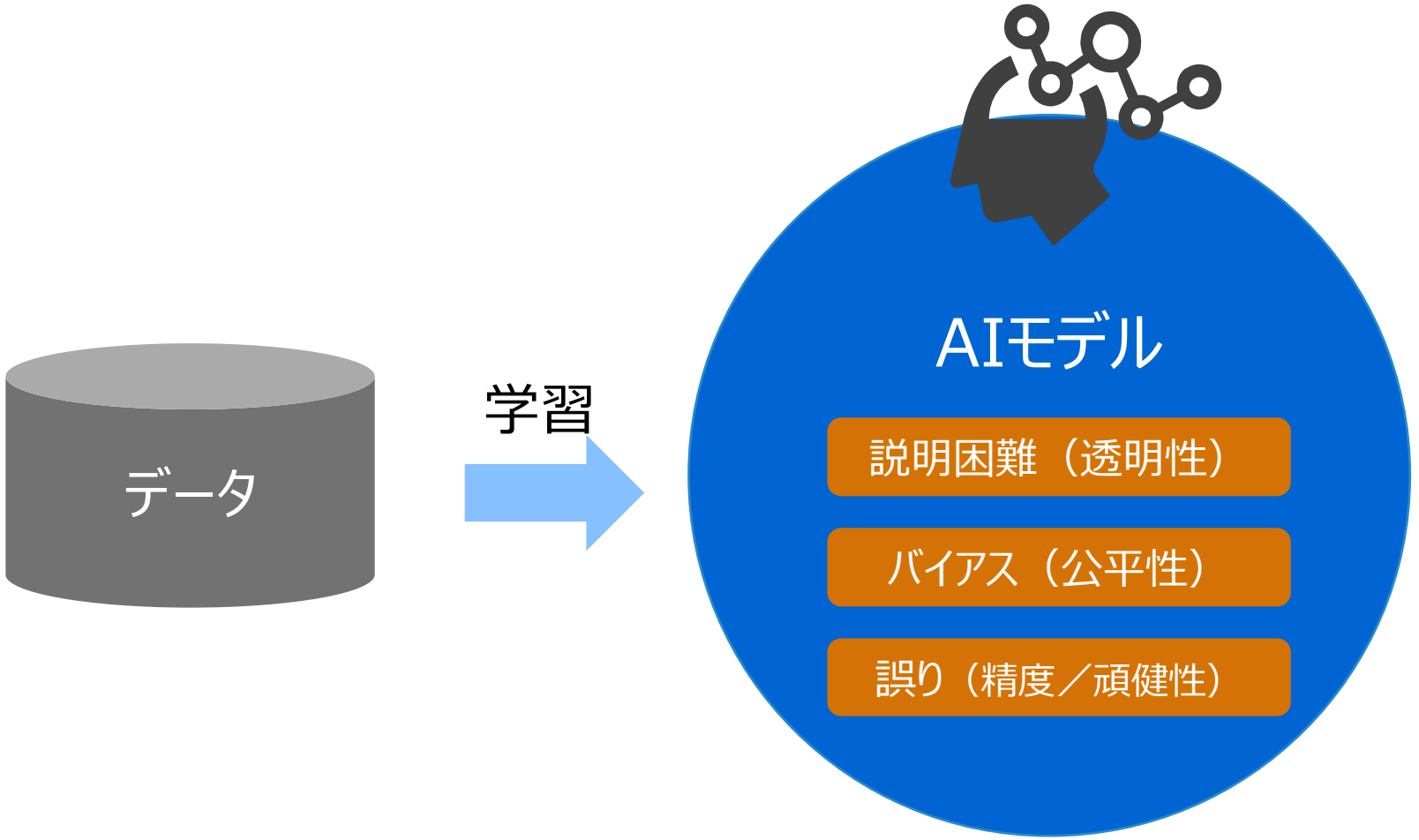
3 誤認識 フェールセーフ（精度・頑強性）



人を動物に誤認識して人権問題に

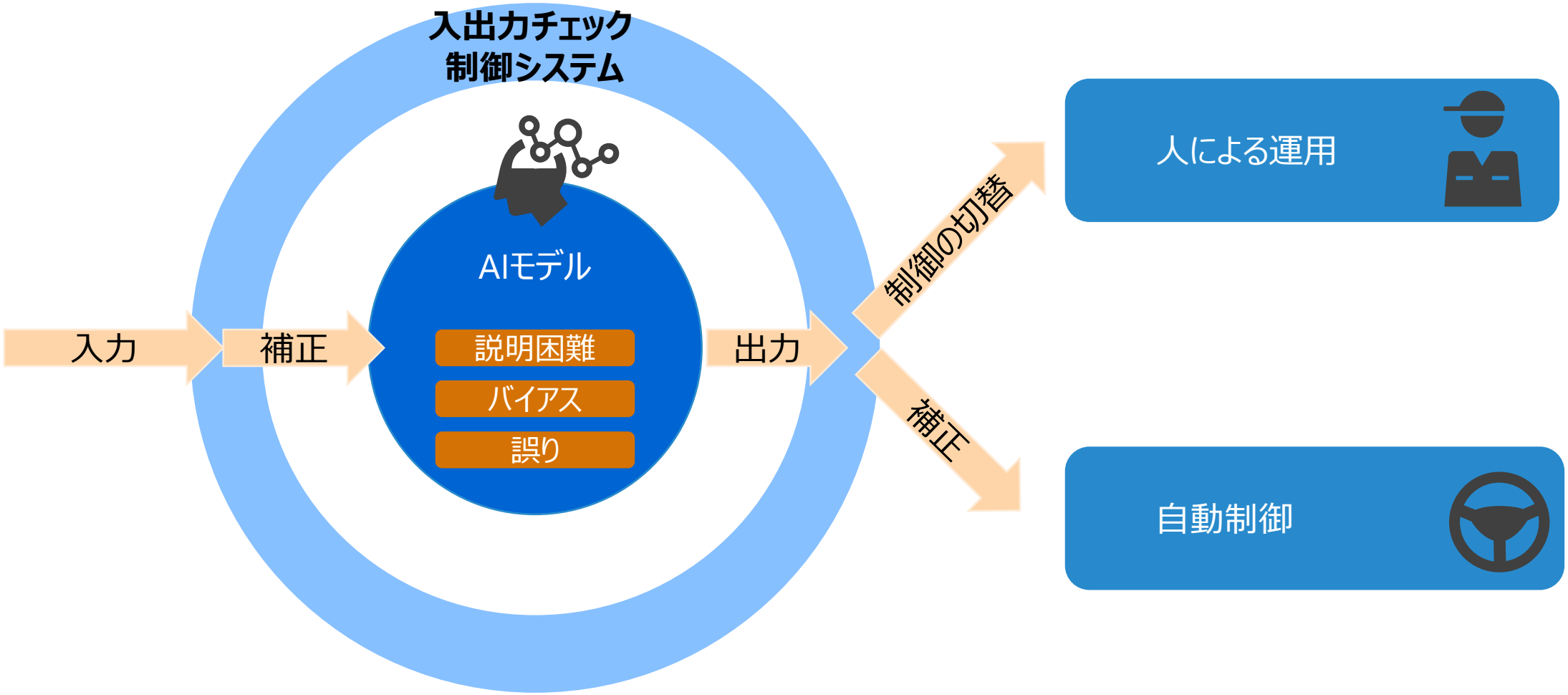
学習データ品質とAIモデルの品質

AIモデルの品質は学習データの品質に大きく影響を受ける
AIモデルの品質は精度の他に特質（透明性・公平性・精度／頑健性等）がある



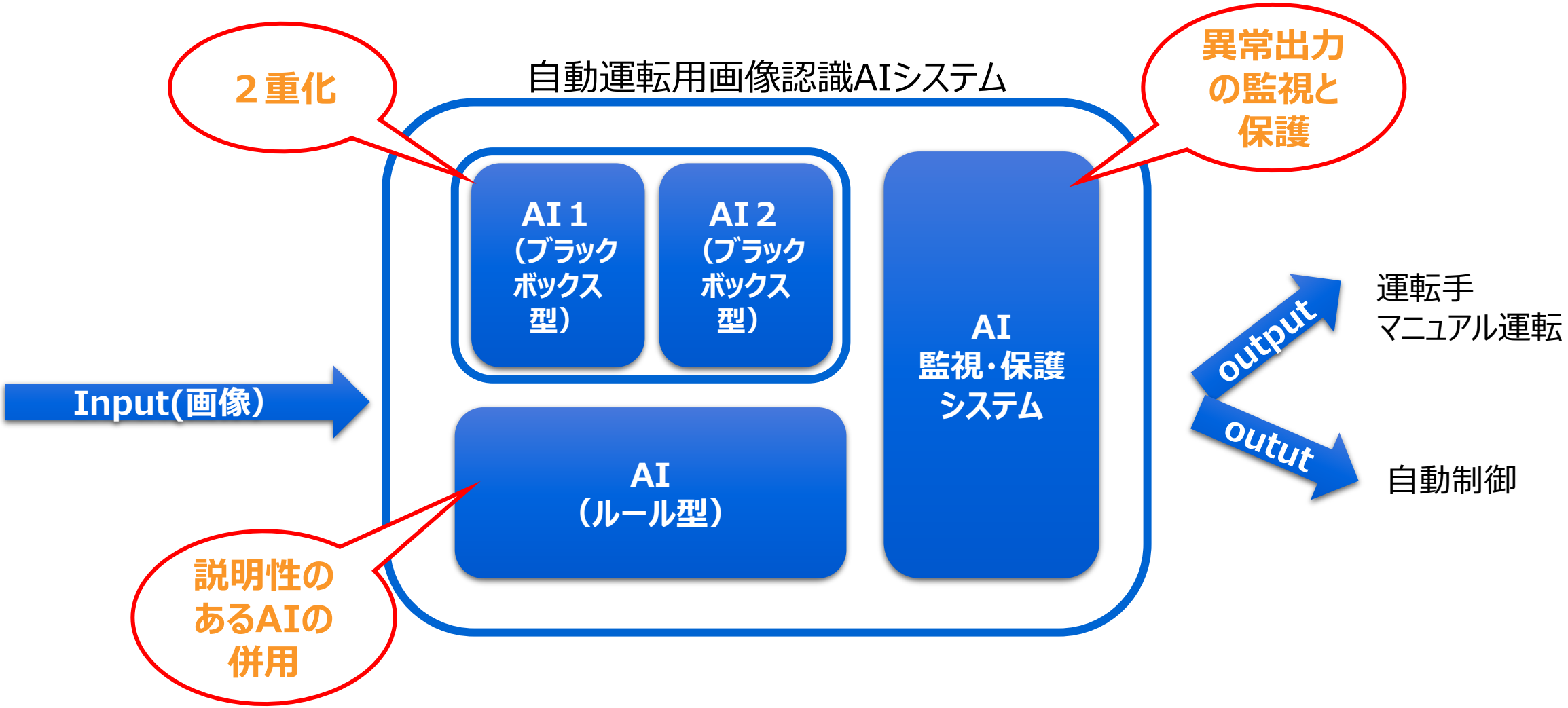
AI搭載システム（AIシステム）と運用の品質

AIモデルの不完全性を補完する“システムと適切な運用”が必要



不確実性のあるAIを使う工夫（自動運転用画像認識モジュールの例）

ブラックボックス型のAIは2重化し見張りをつける



メンテナンス業務における事例紹介①（発電プラント異常予兆検知AI）

実証実験中：
株式会社シグマパワー有明
三川発電所（福岡県大牟田市）

数百～数千の
センサデータ

異常予兆検知AI

アラート

AIの不確実性
を人間系でカバー
（判断結果）

センサー値の微小な変動を分析し、
人よりも早期に異常の兆候を検知

保守員がメンテナンス
の必要性を判断

AIモデル開発

特定期間のデータを
用いてAIを構築

運用時

運転中に、正常値が変化

センサ値

正常の状態が変化

⇒ AIの性能が低下

メンテナンス時

AIの不確実性
を人間系でカバー
（AI性能）

センサ値
センサ値
センサ値

異常予兆検知AIが性能劣化も検知
⇒ 保守員が確認の上、AI再構築指示

AIモデルの再学習・現場展開

TOSHIBA SPINEX™ for Energy（サービス基盤）

メンテナンス業務における事例紹介②（東芝キヤリア様 冷媒漏えい検知システム）

冷媒漏えいを早期に発見し、冷媒漏えい量を削減

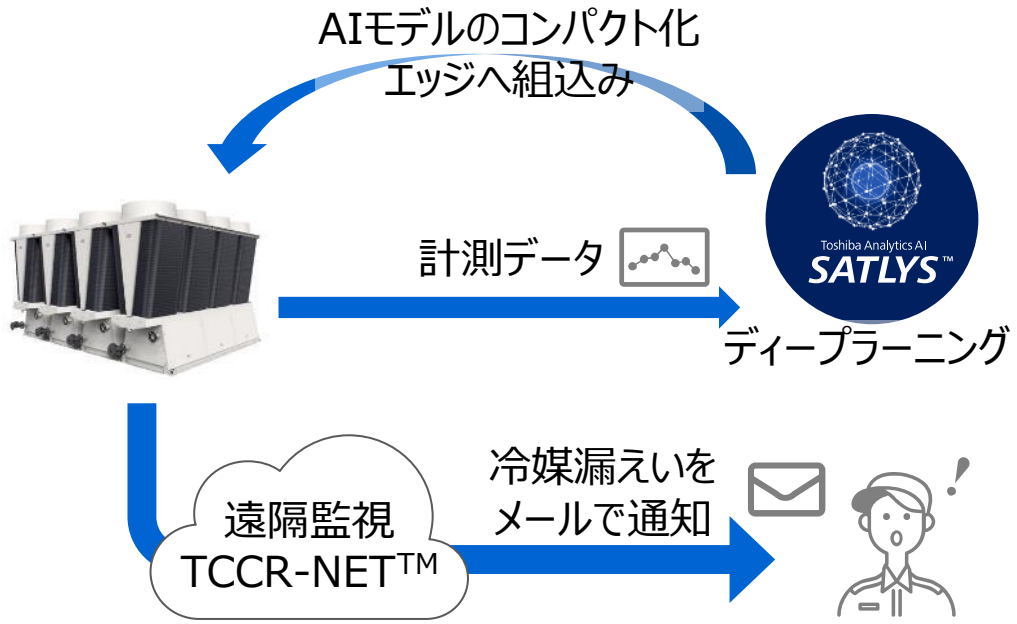
東芝アナリティクスAI SATLYS™

東芝デジタルソリューションズ株式会社

- 冷媒漏洩を早期に発見することにより、万が一の冷媒漏えい発生時に、お客様の熱源機の冷媒漏えい量削減のみならず熱源機の運転効率の低下による消費電力の増加や能力不足の抑制に貢献
- JRA GL-17[*1]対応を実現（業界初[*2]）

熱源機の遠隔監視における
現場の課題

- ✓ 冷凍サイクルの冷媒漏えいをデータから検知する術がなかった
- ✓ 温室効果ガスの漏えい
- ✓ 重故障への発展による保守費の増加
- ✓ 長期停止による顧客満足度の低下



故障で機器が停止するよりも前にAIが冷媒漏洩を検知

*1 業務用冷凍空調機器の常時監視によるフロン類の漏えい検知システムガイドライン
*2 2021年12月現在。空冷ヒートポンプ式熱源機(空冷式チラー)において。東芝キヤリア(株)調べ
【参考】東芝キヤリア(株)ニュースリリース https://www.toshiba-carrier.co.jp/news/press/220126/
【参考】TOSHIBA SPINEX Marketplace https://www.spinex-marketplace.toshiba/ja/services/tccr-net

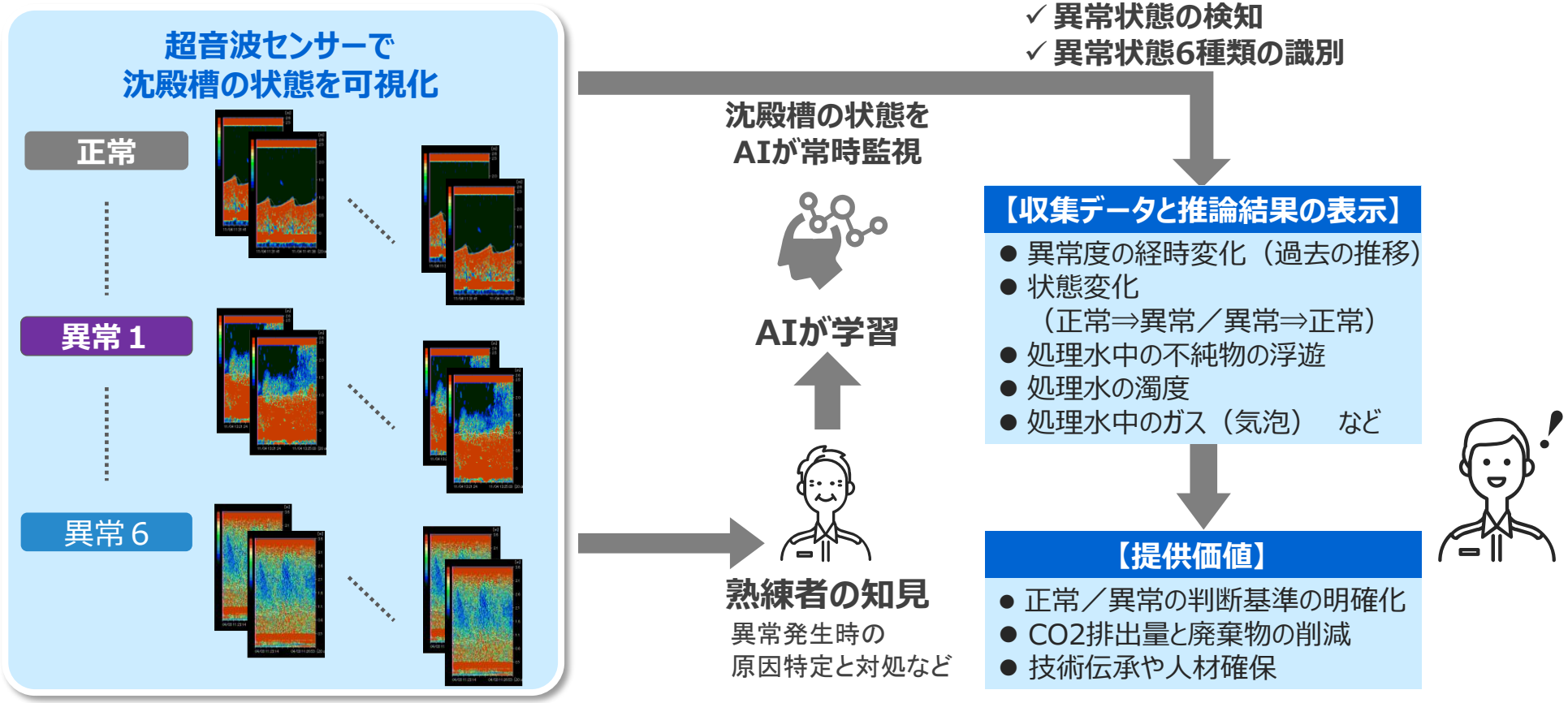
メンテナンス業務における事例紹介③（栗田工業様 水処理設備 沈殿槽自動監視サービス）

1.異常検知AIモデル

沈殿槽内の異常状態を高精度で検知（正常時のデータのみを用いて、教師なし学習で構築）

2.状態推定AIモデル

沈殿槽で複数種類の異常状態の高精度な識別（正常時のデータ、および複数種類の異常状態のデータを用いて教師あり学習で構築）



出典：https://www.kurita.co.jp/aboutus/press220804.html
出典：https://www.global.toshiba/jp/company/digitalsolution/news/2022/0804.html

AI品質保証

1. AI品質は、顧客の期待値に応えるAIシステムの性能によって決まる

2. AIシステムの性能を支える要素は、
“学習データ・AIモデル・制御システム・運用”である

3. AI品質保証とは、顧客の期待値に応えるために、
AIシステムの性能を保つことである

4. すなわち、“学習データ・AIモデル・制御システム・運用”の
品質を全体で保証することである

AI品質保証とは、顧客の期待値に応えるための性能の要素である
“学習データ・AIモデル・制御システム・運用”の品質を全体で保証すること

03

AI品質保証の詳細

東芝の考えるAI品質保証　企画（プランニング）・開発・お客様の運用まで全体をカバー

東芝のAI品質保証の全体像

AI品質保証のガイドラインを起点に、チェックプロセス・品質評価・提示へ

AIシステム品質保証ガイドライン
品質保証の観点・指針

AIシステムの品質保証プロセス
AIビジネス検討 → AIモデル開発 → AIシステム開発 → AIシステム運用

開発における品質保証の進め方・チェックリスト

AIテスト技術
テスト対象 × 品質特性

テストの観点を網羅

具体化

品質評価技術
AIの品質特性を評価する技術

- モデル品質（頑健性など）
- データ品質（被覆性など）

AIシステム品質の文書化・共有
品質カード

1．品質保証の指針となるガイドライン
AIシステムをどのように品質保証するのか？
品質保証の観点や指針を示す

2．品質保証のためのプロセス
開発の中で品質保証をどのように進めるのか？
進め方の手順とチェックリストを整備

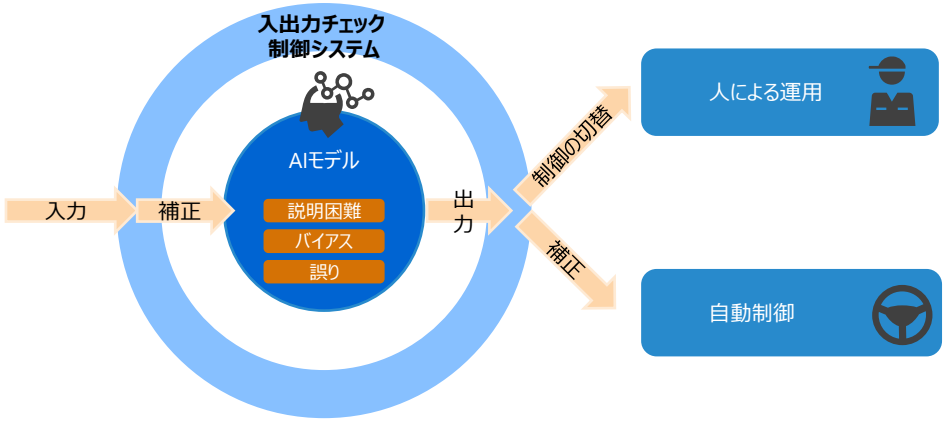
3．AIモデルの品質評価技術
データ・モデル・システムをどう評価するのか？
網羅的な評価の観点や具体的な技術を開発

4．顧客運用のため品質提示
顧客や利用者にAIシステムの品質をどう示すか？
品質カードを使った提示

AI品質保証の指針となるガイドライン

AIを搭載した製品開発で「何に気をつけなければならないか」を整理

品質保証の考え方
何をすべきか
“What“



AIシステム

品質保証の５軸
データ
モデル
システム
開発プロセス
顧客

✖

ステークホルダー
AIプランニングチーム
AIアナリストチーム
AIシステム開発チーム
AIシステム運用チーム
品質保証チーム

✖

開発工程
AIビジネス検討
AIモデル開発
AIシステム開発
AIシステム運用

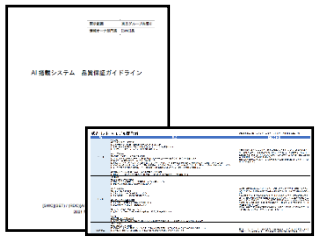
AIプロダクト品質保証ガイドラインを参考

観点を
網羅的に記載

AI品質保証のためのチェックプロセス

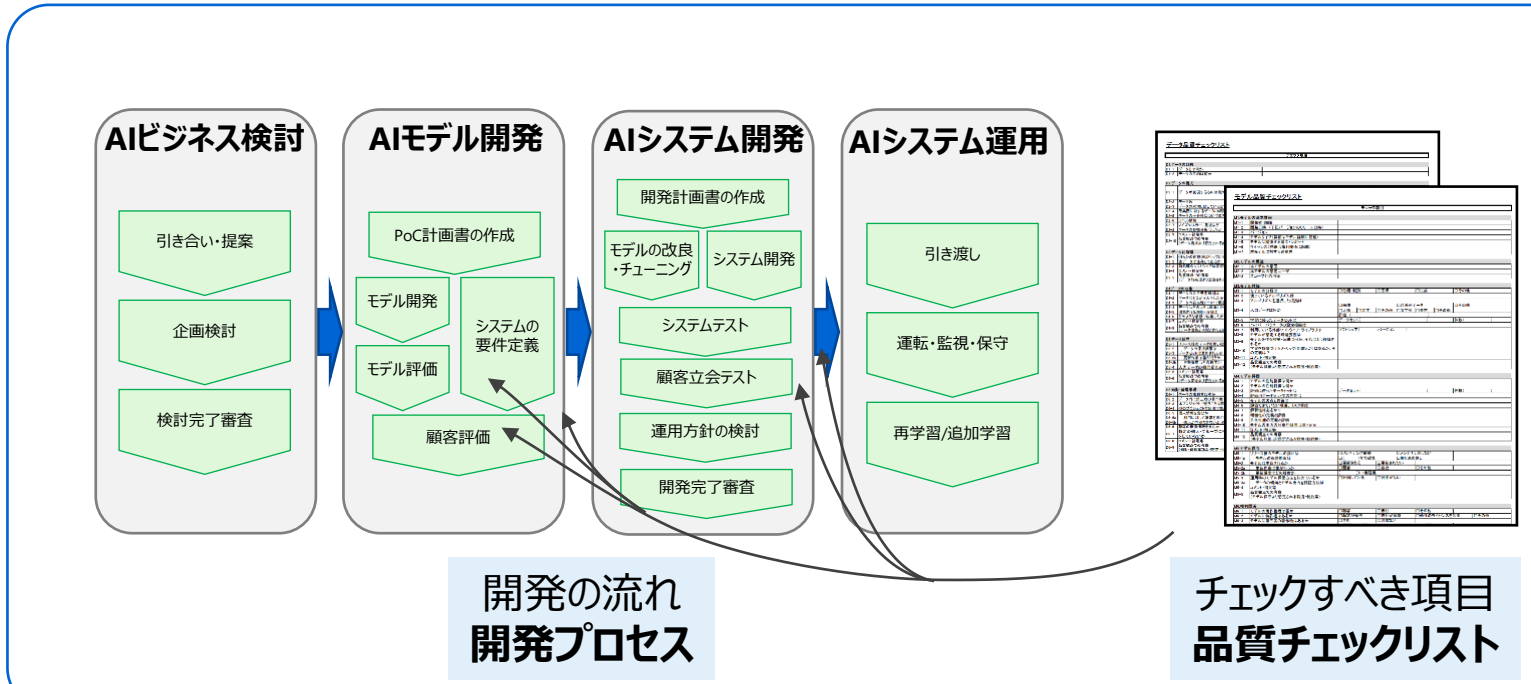
ガイドラインに沿った開発を実現するために、開発の手順や作成すべき成果物を整理

必要な作業、作成すべき成果物開発中に実施できるように手順やチェックリストを定義



AIシステム
品質保証ガイドライン

開発の中で
どこで、どのように
実施するか



AI品質保証プロセス

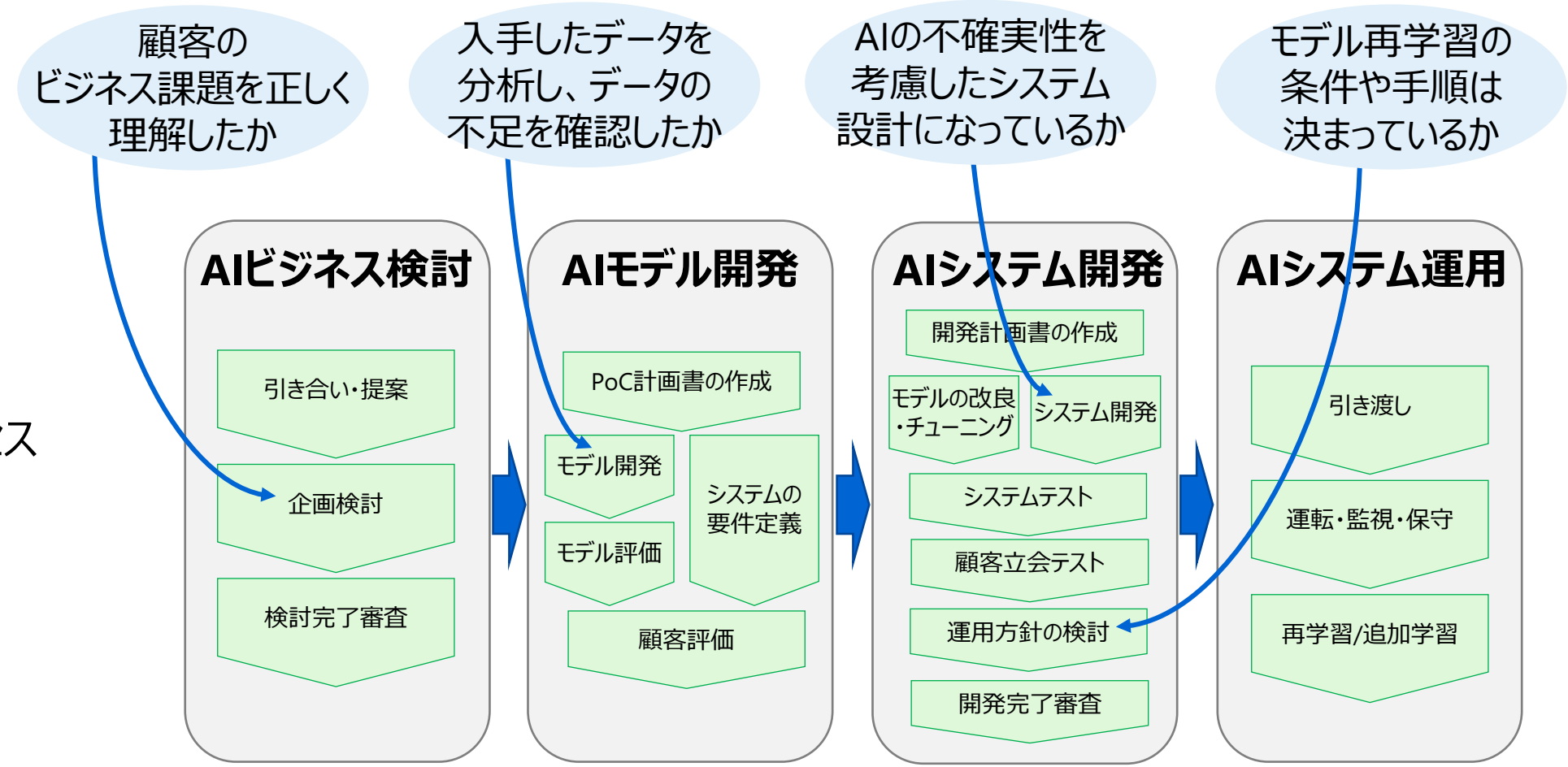
ガイドラインと開発プロセスの対応付けの例

AIシステムの開発プロセスにガイドラインを対応づけることで、観点を網羅

ガイドラインの観点
（やるべきこと）

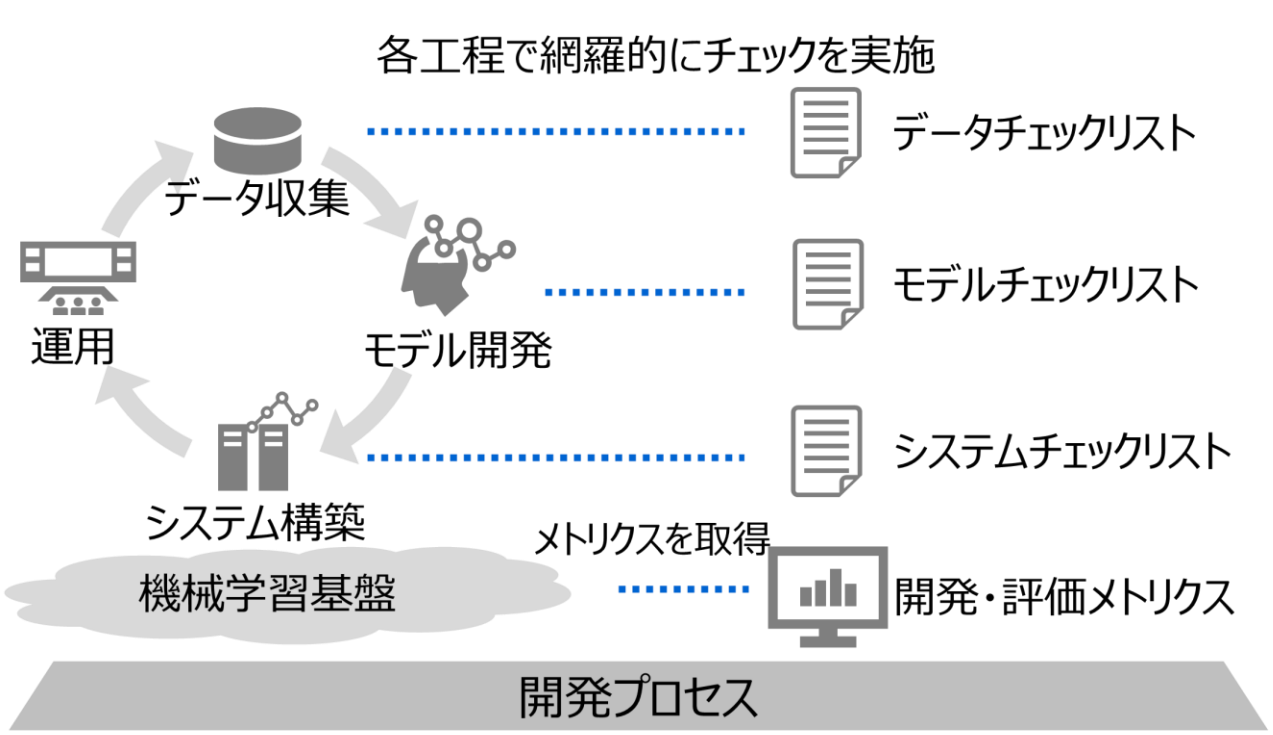
対応
づける

AIシステムの開発プロセス
（どこで実現するか）



適切な運用のための品質カード

システムの開発中に確認したチェックリストや収集した各メトリクスから、要点をまとめて品質カードを作成し顧客に提示（顧客運用の注意点整理）



開発プロセスに従い、開発目線で品質管理・品質保証

利用時に重要なポイントを整理し、提示

データ品質カード
- 学習に使ったデータ
- 学習していないデータ
- データの分類・傾向

モデル品質カード
- モデルの評価結果
- 検証できていない条件・環境
- モデルの強みと弱み

システム品質カード
- システムの期待できる性能
- 検証できていない条件・環境
- モデルの強みと弱み

顧客・利用者向けにシステム品質の提示・理解

品質カードの例

システム品質カード

システム目的

プラントにおける自動制御システム
・自律制御を行うことで、正常運転しつつ最もエネルギー効率を高める
・季節によらず、適切な制御とエネルギーコスト低減を狙う

システムリスク

(1) 制御量が不適切の場合、プラントの出力異常を引き起こす
制御量は多すぎても少なすぎても出力異常を引き起こす可能性がある
主な出力異常の原因は、センサAがXXXを超えている場合や
センサBがXXX以下の場合に制御を行わない場合である
(2) 制御回数が過剰の場合、コスト増を引き起こす
制御１回にごとに動作のためのエネルギーコストがかかるため、
制御回数が多くなるとコスト増となる

テストデータによるシステム性能

テストデータ（2020/7～2021/6)による
・エネルギー使用量(対手動制御) 103 %
（コスト換算：手動と比較して約200万円の悪化）
・網羅性評価
- 手動制御と比較した差が最大の月 8月 差5.8%
- テストデータ全期間の手動-AIの差の変動（右図）

テストデータによるシステム性能（続き）

・自動制御による出力異常の発生回数 2回
- 出力異常の原因：1回目 8/17 (盆明け月曜日)、2回目 1/4 (年始
・頑健性評価
- センサにノイズを載せて評価：結果 エネルギー使用量の変動は
- センサデータが取得できない想定で評価：10分後にアラート発生(

検証できていない項目

本システムの学習データは2018/8～2020/7である
この期間に発生していない以下の条件は検証できていない
・気象現象：大雪、冷夏、・・・
・社会現象：新型コロナウィルスにおける行動変容

運用方針

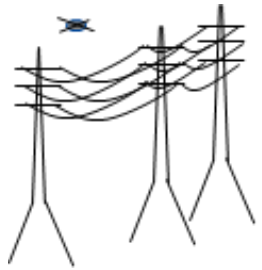
休暇明けに出力異常が発生する可能性が高いため、通常は自動制御
休暇明け日はオペレータによる常時監視を必須とする
夏季の精度が悪いため、夏季はオペレータによる運用状況確認を継続

システムの目的
※顧客が求めるものは何か
↓
システムリスク
※何が起こる可能性があるか
↓
テストデータによるシステム性能
※実際にシステムの性能はどうか
どのようなときにうまくいかないか
↓
検証できていない項目
※確認できていないデータや
条件はあるか
↓
運用方針
※どのようにシステムを活用
することが推奨されるか

※注：本システム品質カードはサンプルとして作成したもので、実際のシステムの品質を示したものではない
品質カードに載せるべき項目は、利用者と共有する目的に応じて選択する

リスクと各フェーズにおける実施事項の例①

ユースケース(※): 送電線の点検において、ドローンが撮影した画像をAIを活用して解析し
保守員が点検すべき異常個所を検出

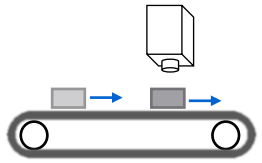


想定リスク	AIビジネス検討	AIモデル開発	AIシステム開発	AIシステム運用
	各フェーズでの対策実施事項			
点検箇所の過剰な識別による点検コストの増大	・顧客との点検レベルの擦り合わせ	・必要な予測精度を有することの評価	・指摘が通常より膨大になった際のアラート通知	・最終的な点検箇所は保守チームで判断 ・過剰に検出された際に再学習を実施
外部環境の変化により識別精度が著しく低下	・送電線エリア・季節などのモデル作成単位の検討	・充分な頑健性を有することの評価	・可用性向上のため複数AIモデルを用いた予測	・各モデルの性能を定期的に評価 ・必要に応じて再学習を実施

(※)東大未来ビジョン研究センター AIサービスとリスクコーデネーション研究会 リスクチェーンモデルのユースケースからリスクと対策を引用して、各フェーズでの実施事項として整理
https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/01/RCModel_Case03_Power-Line-Inspection-AI_JP.pdf

リスクと各フェーズにおける実施事項の例②

ユースケース(※): 製造プロセスの中で画像診断による検品AIを組み込み正常品、異常品を分配



想定リスク	AIビジネス検討	AIモデル開発	AIシステム開発	AIシステム運用
	各フェーズでの対策実施事項			
製品の仕様変更により、識別精度が著しく低下	・製品仕様変更時の再学習等の取り決め	・精度、汎化性能、頑健性の評価 ・モデル開発時評価結果の記録保存		・製品仕様変更時の再学習などのプロセス定義と実施
画像にノイズが混入しAIの判断精度が劣化		・充分な頑健性を有することの評価	・カメラの要求仕様明確化と実装 ・AIの判断履歴の記録保存	・撮影環境の清掃・維持

(※)東大未来ビジョン研究センター AIサービスとリスクコーデネーション研究会 リスクチェーンモデルのユースケースからリスクと対策を引用して、各フェーズでの実施事項として整理
https://ifi.u-tokyo.ac.jp/wp/wp-content/uploads/2022/01/RCModel_Case04_Defect-Detection-AI_JP.pdf

04

MLOpsによる継続的な品質チェック

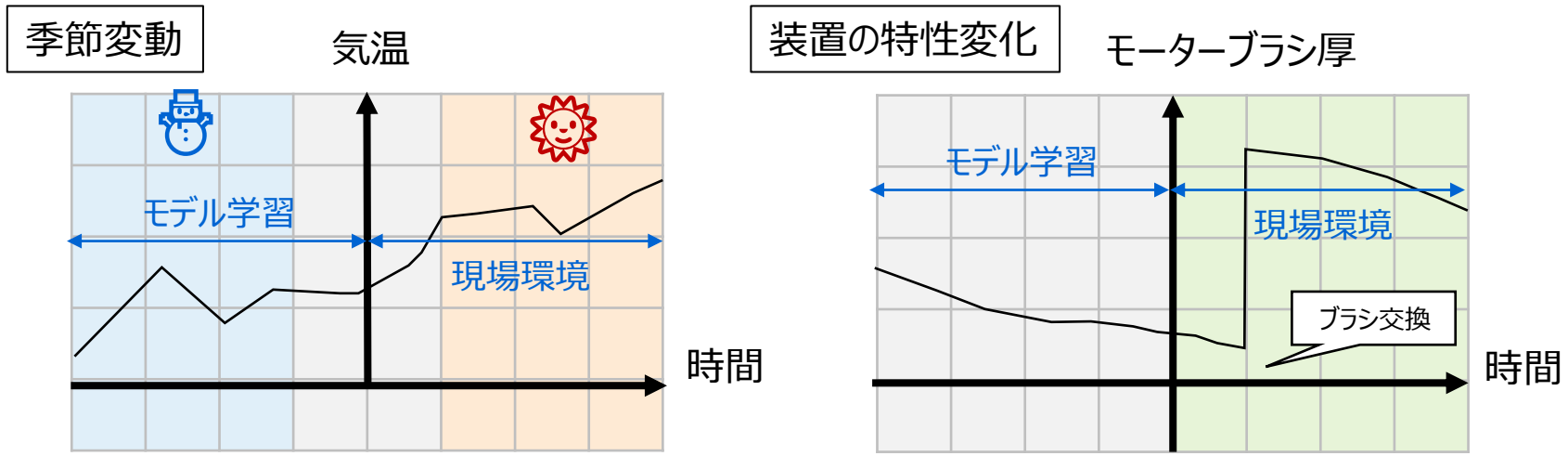
運用時の性能モニタリング

運用では機械学習のモデルの動作状況を監視し、精度低下の検知が必要

モニタリングの必要性

- 運用ではデータの性質の変化によりモデルの性能が低下する
- 特に注意が必要なもの
  - **精度の低下**（画像分類の予測精度など）
    - 偽NGの増加に伴い救い上げ工数が増加し、ビジネスKPIに影響する可能性

精度低下の要因例）



モニタリングは現場環境に精通し、精度の低下要因を考察できる人が担当することが望ましい

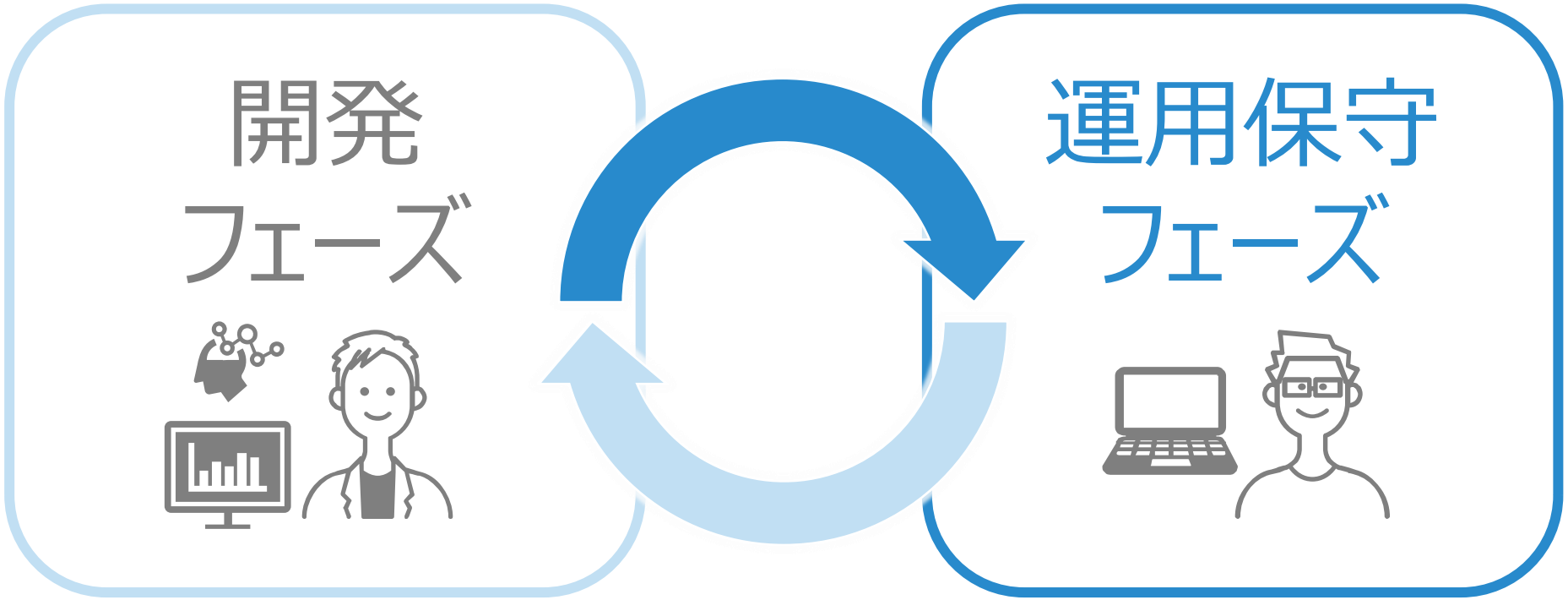
▶モデルを学習したデータと現場の環境と異なると精度が低下する

MLOpsとは

機械学習の専門家・ビジネスの専門家・運用の専門家・システム開発の専門家が
一体のチームとなってサービスを継続的に性能チェックし改善していく取り組みと仕組

Machine Learning + Development + OperationS

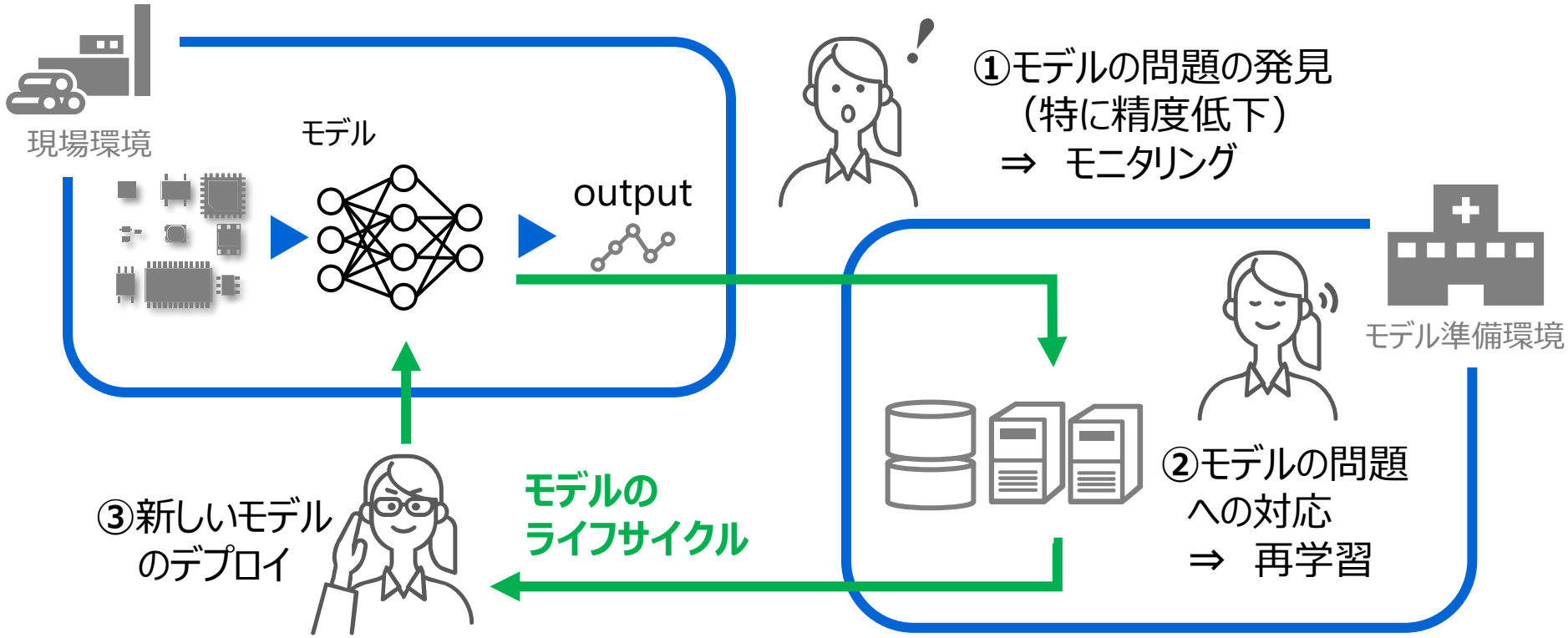
機械学習　ITシステム開発　システム運用



MLOpsの「Ops」（運用）が目指す姿

**Opsチームはビジネス KPIの達成に貢献するため、
モデルの問題を発見し、その対応を行う**

AIモデルの保守は、AIモデルの再学習（作り直し）を伴う



まとめ

1. AI品質保証とは、顧客の期待値に応えるための性能の要素である“学習データ・AIモデル・制御システム・運用”の品質を**全体で保証**すること

2. AI品質保証ガイドラインで**チェック観点を網羅的**にリスト化

3. 個々のAI製品の開発プロセスに観点リストを対応づけ、**誰が・いつ・どこで・何を・どのようにチェックするか整理**し、チェックリストを作成

4. チェックリストを”データ・モデル・システム・運用”ごとにお客様がリスクを理解し**適切な運用ができるように品質カードを作成し提示**

5. MLOpsで**運用時にAIシステムの性能モニタリング**が必要な場合がある


TOSHIBA

AI品質保証にまつわる国内外の動向

日本の動向

政府・国機関
内閣府・総務省・経済産業省
産業技術総合研究所（AIST）
科学技術振興機構（JST）
⇒各種指針・ガイドラインの公開

学会・研究会
機械学習工学研究会（MLSE）
AIプロダクト品質保証コンソーシアム（QA4AI）
⇒品質保証やエンジニアリングの研究・議論
勉強会の開催、ガイドラインの作成

企業
メーカー、IT企業　etc.
⇒AIの指針や品質保証の仕組みの整備と展開

国際社会の動向

国際標準化
ISO/IEC JTC 1/SC 42 (Artificial Intelligence)
UL 4600（自動運転の安全評価規格）
⇒国際標準の策定

法規化・ルール
EU AI規制法案
⇒強制力を持つ法規の検討
アメリカ NIST
⇒AI利用時のリスクマネジメント

企業
IT企業 etc.
⇒独自の指針や研究成果のリリース

AIシステム品質保証に関わるガイドライン・EU法案の比較

	AIプロダクト品質保証ガイドライン（QA4AIガイドライン）	機械学習マネジメントガイドライン（AIQMガイドライン）	欧州 AI規制法案（EU AI act.）
作成団体	AIプロダクト品質保証コンソーシアム（産学の任意団体）	産業技術総合研究所（メンバには産学も含む）	EU Committee
リリース日	2019/5(初版)、 2022/7(最新版)	2020/9(初版)、2021/7(最新版)	2021/4(法案提出)、2024施行(予定)
目的	主に、AIプロダクトの**開発者や品質保証に関わる人向けに、品質保証の共通的な指針**を与える	機械学習システムや機械学習要素に関する**品質の基準と達成目標**を定めることで、企業が品質を測定し、AIの誤りに起因する損失を減少させる	**基本的権利の保護とユーザーの安全**を確保することと、AIの開発と普及に対する信頼を強化すること
内容	**品質保証の5軸と観点** 品質保証に応用可能な技術カタログ **製品ドメインごとの解釈**・応用	利用時品質・外部品質・内部品質といった**体系的な品質の考え方** 品質レベルに応じた内部品質の要求事項	**AIシステムをリスクに応じて4つに分類**し、分類ごとに遵守すべき内容を定義 特に「ハイリスク」に分類されるシステムは法規への対応や監査を義務づけ
ポイント	• 学術界よりも産業界の意見が多く含まれている • 体系的にまとめるよりもノウハウを蓄積して具体的な内容にまとめている • 年１度程度の更新が継続している	• 体系的な品質マネジメントシステムを構築できるようまとめている • 品質レベルの考え方を導入し、内部特性の要求事項を定めている • 国際標準化への提案を進めている • 別のガイドラインへ派生している (プラント保安分野AI信頼性評価ガイドライン) • 品質評価のためのツールも開発中	• 法規として、域外適用も含めてAIを活用する際に遵守が求められる • 制裁金の導入も予定されている • **リスクベースの考え方**であり、リスクの評価とその対応が強く求められる • AIの精度面よりも安全面や倫理面に関する要求が強い

AIシステムを実行するときの規制上のリスク（欧州AI規制法案:2024年に施行予定）

リスクに応じて4段階に分類。罰則は3,000万€か世界年間総売上高6%の高い方