特集
Special Feature

［AI の品質保証］

2 機械学習品質マネジメントの体系化に向けて

応用 一般



小西弘一　大岩　寛　妹尾義樹

産業技術総合研究所

AI への期待が高まって久しい．自動運転や金融与信審査など人の生死や人生を左右する場面での利用も期待される中，その役割に見合う品質が必要である．このような認識が，ここ数年で広まり，世界各地で AI が備えるべき品質やその実現手法を示す文書の整備が始まっている．日本は世界に先駆けて AI 品質マネジメント手法の体系化を進めている．本稿ではその取り組みの概要を紹介する．

機械学習品質管理の課題

ICT は生活に浸透した．公共インフラにもコンシューマー機器にもソフトウェアによる高度な制御が使われている．中には，航空機や医療機器のように人命にかかわる機器や，人材採用や金融与信審査など生活基盤にかかわる判断を行うものもある．これからの ICT に求められる機能は複雑で，そもそも仕様を適切に定めることさえ困難なことがある．AI，中でも機械学習に基づくソフトウェア構築手法は，そのような状況を打破するのに役立つ．そのため，ソフトウェア構築における機械学習 AI への依存度が増え続けている．

機械学習 AI 品質管理の難しさとその影響

機械学習 AI を取り込んだソフトウェアの品質管理には，後述の理由により，従来の手法をそのまま適用することが難しい．

従来のソフトウェア品質管理との違い

従来のソフトウェアの品質管理手法は確立しており，日本が大きな役割を果たす形で国際標準化（SQuaRE ISO/IEC25000 シリーズ）も進んでいる．従来のソフトウェアの品質管理の基本的な考え方は，ソフトウェアが満たすべき仕様を人が段階的に詳細化して細部まで明確にし，そのすべてを細部から全体に戻る順にテストして仕様の充足を確認する，というものである．

ところが，機械学習 AI の利点は，人が仕様を定めきれないほど複雑なソフトウェアを，人が定める仕様の代わりに大量のデータを用いて合成できることにある．このため，機械学習 AI を取り込んだソフトウェアの開発では，仕様に基づくテストを行えない個所が生じ，従来手法では品質を保証できない．

ビジネスへの影響

機械学習 AI の品質管理手法が明らかでないと，ビジネス上，以下のようなさまざまな問題が生じる．

1）AI を適用したソフトウェアを顧客に安心して使ってもらえない．

機械学習 AI を取り込んだソフトウェアを開発した企業やそれに基づく製品やサービスを提供する企業は，そのソフトウェアが適切な品質管理の下で開発されたことを顧客に説明できない．その結果，その製品やサービスの利用者は，そのソフトウェアを安心して利用できない．

2）問題が起きたとき，説明責任を果たせない．

結果的に不十分な品質のソフトウェアが出荷され，事故を引き起こしたとき，そのソフトウェアを開発した企業は，品質実現のために行った努力の妥当性を説明できない．その結果，過度な製造物責任を問われる可能性がある．

3）高い品質を実現しても，それに基づく差別化ができない．

他社より優れた品質を実現しても，他社のソフトウェアとの品質の違いを顧客に理解してもらえない．その結果，品質が自社の製品やサービスの魅力増大に寄与しない．

機械学習 AI を取り込んだソフトウェアを開発する企業は，上記の問題に気付いている．しかし，この問題を解決するには各社個別の取り組みだけでは十分でない．機械学習 AI の品質管理に関する社会共通の理解が必要である．そこで，共通理解を構築する取り組みが世界で進められている．

世界各地での取り組み

機械学習 AI の品質管理の難しさが社会に認識されるにつれ，世界各地で，機械学習 AI が備えるべき品質やその管理手法の基準作りが始まった．

NIST が主導する米国

米国では National Institute of Standards and Technology（NIST, 国立標準技術研究所）が 2021 年 7 月に AI Risk Management Framework[☆1] の開発計画を公表し，コメント公募やワークショップ開催により世界中の意見を集めながら検討を進めている．2022 年 3 月には最初のドラフトを発行した．AI に基づくシステムのリスク，備えるべき品質，およびその実現管理手法を示している．第 1 版の発行は 2022 年末から 2023 年初めの予定である．将来，このフレームワークが米国政府調達要件になると，サプライチェーンに連なる日本企業にも影響が生じると思われる．

AI 法案を準備する欧州

欧州は欧州委員会が主導して法規制を目指しており，2021 年 4 月に法案（通称 The AI Act[☆2]）を公表した．人権侵害リスクが高い，いくつかの特定用途での AI 利用を禁止し，また人命や人権への影響が大きい用途での AI 利用について要件を定め，認証対象としている．当然，欧州市場を対象とする日本企業にも遵守が求められる．

国内の省庁やアカデミアの動き

国内では，ガバナンスから品質管理への技術アプローチまで，さまざまな活動が進んでいる．内閣府は 2019 年に「人間中心の AI 社会原則」[☆3] を作成した．経済産業省は 2021 年 7 月に「AI 原則実践のためのガバナンス・ガイドライン」[☆4] を発行した．AI プロダクト品質保証コンソーシアム（QA4AI）は 2019 年 5 月に「AI プロダクト品質保証ガイドライン」[☆5] を発行した．日本ソフトウェア科学会機械学習工学研究会（MLSE）は 2022 年 6 月に「機械学習システムセキュリティガイドライン」[☆6] を発行した．

機械学習品質管理手法の開発

国立研究開発法人産業技術総合研究所（以下，産総研）は，世界各地の動きに先駆けて，2018 年度から機械学習 AI の品質保証の取り組みを開始した．国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）からの受託事業（JPNP20006）の一環として，1 年間の先導研究を経て，2019 年度から 2023 年度までの実証研究を実施中である．有識者会議（後述）を組織し，産業界と連携して活動している．

☆1 https://www.nist.gov/itl/ai-risk-management-framework

☆2 https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

☆3 https://www8.cao.go.jp/cstp/ai/aigensoku.pdf

☆4 https://www.meti.go.jp/press/2021/01/20220125001/20220124003.html

☆5 https://www.qa4ai.jp/

☆6 https://github.com/mlse-jssst/security-guideline

日本発の品質管理を世界へ

本活動の狙いは AI 品質管理の課題を解決して，機械学習 AI のビジネス活用を促進することにある．また，日本の AI の強みを作ることも念頭に置いている．AI 開発で米国や中国と伍して戦うのは難しい一方で，伝統的に日本が得意な品質管理ならば AI においても日本が主導し，優位性を築くチャンスがあると考えている．

原則から実践まで幅広い取り組み

本活動は，社会原則や法規制を踏まえて組織ごとのガバナンスにより品質目標が決まった後，目指す品質を実現するための原則から実践を支援するまでを検討対象としている．

具体的には以下の 4 つの成果目標を置いた．

- ガイドラインの開発
- リファレンスガイドの開発
- 品質マネジメント支援環境の開発
- 品質評価・向上技術の研究開発

ガイドラインは，AI 品質の考え方を示して，顧客と開発者の意思疎通を助け，顧客の安心と品質による差別化の実現に寄与する．リファレンスガイド，品質マネジメント支援環境，品質評価・向上技術は，品質の実現自体に寄与し，開発者が顧客に対する説明責任を果たせるようにする．

ガイドラインの開発

機械学習品質マネジメントガイドライン[☆7] は，機械学習 AI を取り込んだソフトウェアの品質の考え方やその実現に必要な事項を，業種によらない分野横断の形で示すものである．社会的要請や事業ニーズを踏まえて品質目標を決める部門と品質保証に当たる部門の間，および，AI を活用したい企業とそこから委託を受けて AI 利用システムを開発する企業の間に，共通理解を提供して，品質目標の合意に役立てることを目標として策定した．

ガイドラインの開発体制

ガイドラインの開発は，産学官から委員を集めて組織した**機械学習品質マネジメント検討委員会**で行っている．民間企業 12 社，国立情報学研究所，東京理科大学，オブザーバとして，(独) 情報処理推進機構，NEDO，経済産業省の委員を迎え，年数回の会議と月数回の詳細検討タスクフォースによる会合を開催している．

機械学習品質マネジメントの考え方

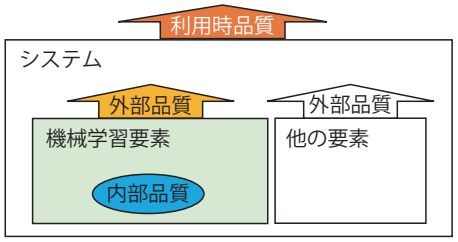
本ガイドラインでは，システム全体で最終的な利用者に提供するべき品質を**利用時品質**と呼ぶ．システムの構成要素のうち，機械学習 AI を取り込んだ要素を**機械学習要素**と呼ぶ．システムの利用時品質を実現するために機械学習要素が備えるべき品質を機械学習要素の**外部品質**と呼び，これを実現することが機械学習品質マネジメントの目標となる．一方，外部品質を実現するために，機械学習要素の設計時や運用時に満たすべき事項を**内部品質**と呼び，外部品質は内部品質を管理することによって間接的に実現されると考える（図 -1）．

本ガイドラインでは品質目標となる外部品質特性として次の 5 つを挙げる．

1) リスク回避性（安全性・危害回避性）
2) AI パフォーマンス（トータルの予測精度）
3) 公平性
4) プライバシー
5) AI セキュリティ

目標とする外部品質のレベルに応じて内部品質ごとの要求事項を示している．

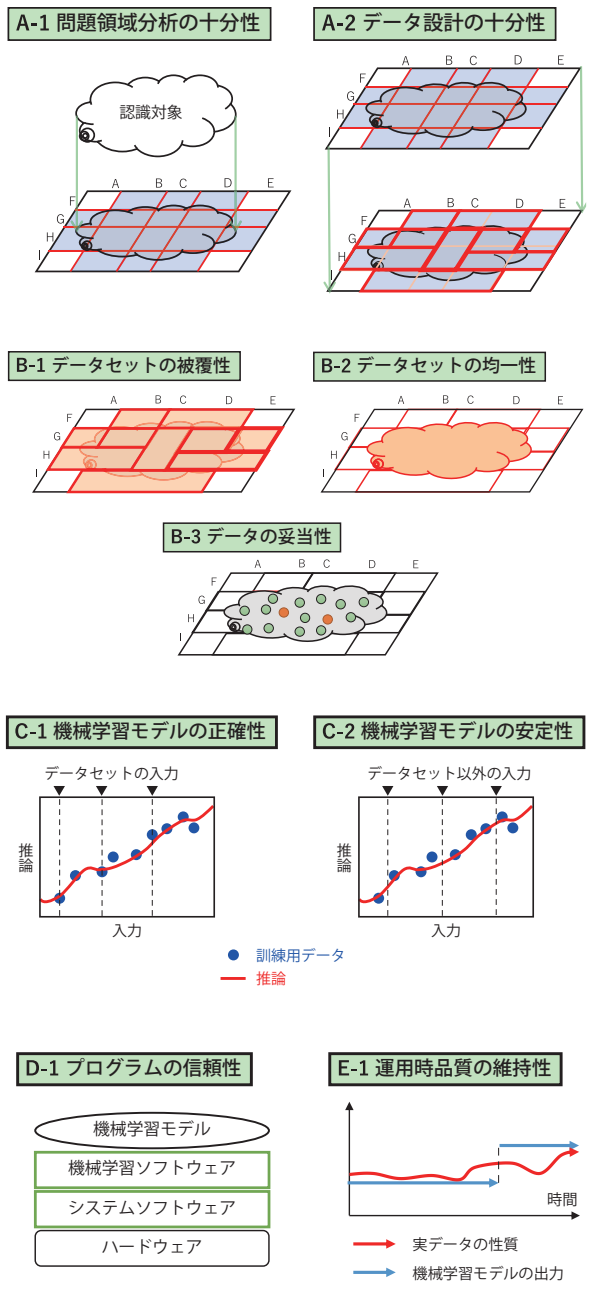
☆7 https://www.digiarc.aist.go.jp/publication/aiqm/



■図 -1 外部品質と内部品質

また品質管理の切り口となる内部品質として以下の9項目を挙げる（図 -2）.

A-1 問題領域分析の十分性：機械学習要素に入力されると想定される運用時の実データの性質についての分析結果が，想定されるすべての利用状況をカバーしていること.



■図 -2　内部品質

A-2 データ設計の十分性：品質管理に用いるデータ整理の細分の枠組みの設計結果が，現実的な枠組み数で実用上十分な網羅性を達成していること.

B-1 データセットの被覆性：細分した領域ごとに十分なデータが与えられていること.

B-2 データセットの均一性：データセット全体の分布が入力されるデータ集合全体の分布に近いこと．被覆性とのバランスが評価の対象となる.

B-3 データの妥当性：1つ1つのデータが妥当なものであること.

C-1 機械学習モデルの正確性：データセットの入力に対して十分に正確な推論が行われること.

C-2 機械学習モデルの安定性：データセット以外の入力に対しても安定した推論が行われること.

D-1 プログラムの信頼性：学習済みモデル以外のソフトウェアの品質が確保されること.

E-1 運用時品質の維持性：運用開始時に確保した品質が，運用期間中を通じて維持されること.

内部品質の実現方法は詳述していない．機械学習の種類や用途ごとに内部品質の適切な実現方法が異なり，また内部品質の実現に用いる品質評価・向上技術が進歩し続けているからである.

リファレンスガイドの開発

ガイドラインは機械学習品質管理の考え方を一般的に示すが，実際に品質マネジメントを行うには具体的手順を示すものがあると便利である．そこで事例や品質アセスメントシートによって手順を示す機械学習品質マネジメントリファレンスガイド[☆8]を開発している.

品質マネジメント事例集

事例集は所与のビジネス要件を満たすための機械学習に基づくシステムについて，機械学習要素の外部品質の要求レベルを見積もり，内部品質の評価や向上を試みた例を集めたものである．現時点では，類型的ビジネス要件を対象に，入手が容易な公開デー

☆8 https://www.digiarc.aist.go.jp/publication/aiqm/referenceguide.html

タを用いた開発例を扱っている．将来的には民間企業の具体的事例を扱って，実際的な品質マネジメントに関する共有知の形成を目指す．すでに参加企業からテーマやデータの提供をいただいての検討が始まっている．

機械学習品質アセスメントシート

リファレンスガイドの一環として，品質アセスメントシートを開発した．製品の機能安全の保証に詳しい企業メンバが主導した．Excel フォームの形で，ガイドラインに沿った品質マネジメントを行う上での検討ポイントのリストと，検討結果の記録のツールを提供している．

品質マネジメント支援環境の開発

機械学習 AI の品質マネジメントを行うには，支援環境も重要である．ガイドラインやリファレンスガイドがあっても，実際にマネジメントを行う際には試行錯誤を繰り返す．機械学習品質マネジメントにかかわる多数の要素の適切な組合せを事前に知るのが難しいからである．そこで，そのような試行錯誤過程を支援する品質評価支援ツール Qunomon[☆9] を開発している．このツールは，品質マネジメントに必要なさまざまな要素を提供しており，また新たな要素を容易に取り入れることができる. さらに品質評価結果からレポートを自動生成する機能を提供する．

とりわけ重要な要素として，機械学習要素の品質を評価する AI テスト技術がある．現状，品質評価技術は発展途上である．したがって，今後優れた品質評価技術が登場し広く普及することで，品質マネジメントの水準を継続的に高めることができる．そこで AI テスト技術の流通を支援し普及を促進するための仕組み AITHub を開発している．AITHub では，AI テスト技術をパッケージ化したものを AIT と呼び，誰もが AIT を登録・検索・入手できるオープンプラットフォームを提供する．

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

品質管理技術およびツールの開発

内部品質の実現に必要な，品質を評価・向上する技術やツールは現在世界的に研究が盛んな分野である．外部の技術を Qunomon 用にパッケージ化する一方，ガイドラインの考え方を踏まえた独自の品質評価・向上技術の開発にも取り組んでいる．

2021 年度は 6 項目の調査・技術開発を行った（**図 -3**）．成果は報告書[☆10] にまとめて公表している．

国際標準化への取り組み

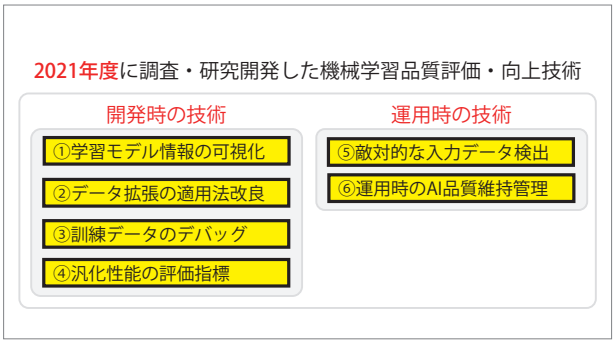
品質マネジメントを日本の AI の強みにするには，日本の品質に対する考え方を世界標準にする必要がある．先に述べた通り，世界各地で AI の品質に関する取り組みは急速に進んでおり，国内標準を確立してから国際標準化の場に持ち込むのでは間に合わない．そこで，ISO/IEC JTC 1 を主な対象に国際標準化を進めている．先行しているのは SC 42（人工知能）における DTR5469 Functional Safety for AI へのインプットである．SC 7（ソフトウェア工学）や SC 27（セキュリティ）の国内委員会とも連携し，オールジャパンでの対応を進めている．

また産総研が持つチャネルを活かし，NIST のフレームワーク担当者とも交流を進めている．

各界からの反響

2020 年 6 月のガイドライン第 1 版の発行以降，さ

☆10 https://www.digiarc.aist.go.jp/publication/aiqm/AIQM-techreport-2.pdf



■図 -3 機械学習品質評価・向上技術

まざまな分野で反響をいただいた．

日本IBMは本ガイドラインに基づいて独自の診断フレームワークを開発し，2021年4月に「ML品質診断サービス」[☆11]を発表した．

経済産業省，厚生労働省，消防庁は共同で2020年11月に「プラント保安分野AI信頼性評価ガイドライン」[☆12]を公表した．機械学習品質マネジメントガイドラインの体系を用いており，プラント保安分野への適用方法を示したリファレンスと位置づけられる．

その他，金融や製造業などの企業や業界団体から問合せを受け情報交換を行っている．

今後の取り組み予定

今後は，成果の社会への普及と適用事例の拡大に向け，以下の取り組みを行っていく．

企業による社会実装の促進・支援

ガイドラインの適用に関心のある企業と連携し，現実的要件に対応するシステムへの適用を進めていく．また，可能な範囲で品質マネジメントの内容と結果をリファレンスガイドの中で公表する．

品質マネジメント手法の拡充

ガイドラインについては5つの外部品質特性をなるべく共通の観点で扱えるよう整理する．

リファレンスガイドでは事例を拡充する．現在の事例は画像処理と分類器に偏っており，すべて教師あり学習である．今後，教師なし学習，時系列データ，自然言語処理などに扱いを広げる．

品質マネジメント支援環境に関しては前述のAITHubサービスの運用開始が重要課題である．加えて，新たなAI品質評価技術とそれに基づくAIT開発・公開を進める．

第三者認証制度

先に述べたビジネス上の課題を解決するには，AI品質に関する共通理解の普及に加えて，企業が行うAI品質マネジメントの妥当性の可視化が必要である．AI利用者に妥当性が評価できるとは限らない．また評価に必要な情報が一般には開示できない場合もある．第三者認証制度があれば，利用者は専門家の評価を参照でき，企業は情報を一般に開示せずに妥当性を訴求できる．評価対象企業に過大な負担をかけることなく，かつ品質実現に有効とするために，制度の在り方について参加企業の知見を集めて検討していく．

先行事例作りへのご参加を！

本活動を意義あるものとするためには，実用的な案件での事例作りが欠かせない．世界に先駆けての事例作りにご協力いただける案件があれば，ご相談いただければ幸いである．

（2022年7月22日受付）

■小西弘一（正会員） k.konishi@aist.go.jp

東大情報工学修士課程修了．IT企業でシステムソフトウェアとセキュリティの研究マネジメントに従事し，2021年より現職．機械学習品質マネジメントプロジェクトの運営に参加し，品質マネジメント事例集の作成を推進．

■大岩　寛（正会員） y.oiwa@aist.go.jp

2005年東京大学大学院情報理工学系研究科 コンピュータ科学専攻博士課程修了．博士（情報理工学）．同年産業技術総合研究所に入所．2021年より，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．ソフトウェアシステム，情報セキュリティ，ネットワークなどの研究に幅広く従事．

■妹尾義樹（正会員） y.seo@aist.go.jp

京都大学工学研究科情報工学専攻修了，博士（工学）．NEC中央研究所にてスーパーコンピュータ研究開発などに従事．産業技術総合研究所で人工知能研究企画室長，標準化推進センター審議役などを経て，現在は標準化オフィサーを務める．

☆11 https://jp.newsroom.ibm.com/2021-04-21-IBM-ML-Quality-Diagnostic-Service

☆12 https://www.meti.go.jp/press/2020/03/20210330002/20210330002.html