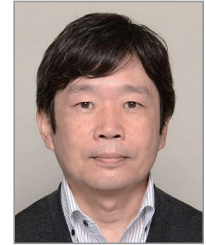


[AIの品質保証]

## 4 SQuaRE 品質モデルの AI へのマッピング



込山俊博 向山 輝 NEC



### AI 応用システムの特徴

情報システムやセンサを通じて蓄積されたビッグデータの利用を背景に、機械学習・深層学習といった技術が進展し、AIを応用したシステムの利用が進みつつある。たとえば、AIによる文章理解や画像認識を組み込んだ応用は自動翻訳やカメラの顔認識など身近なところに及んでいる。

従来型の ICT (Information and Communication Technology) システムの機能実装に AI を利用した AI 応用システムの普及に伴って、開発側にとっても利用する側にとっても、その品質は大きな関心事となっている。その活用が、財産・生命・環境などに影響する領域に広まるにつれ、AI 応用システムの品質の重要性はさらに高まる。

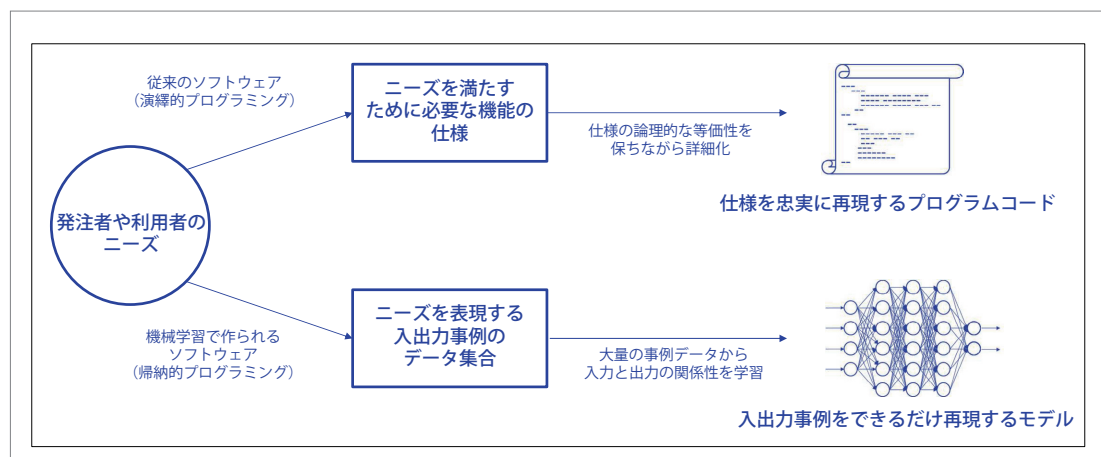
AI 品質を考えるのに先立って、ここでは対象となる

AI 応用システムの特徴を整理する。

従来型の ICT システムの開発は、発注者・利用者のニーズに基づいて、求められる機能や品質を仕様化し、その振舞いを段階的に詳細化し、プログラミングしていた。この方式を、演繹的プログラミングと呼ぶ。一方、AI 応用システム開発の AI を用いた機能実装部分では、入出力の事例としてのデータを多量に与え、事例の中に埋め込まれた仕様を自動的にプログラム化する。この方式を、帰納的プログラミングと呼ぶ<sup>1)</sup>。それらの違いを図-1に示す。

帰納的プログラミングを含む、AI 応用システムの開発において、次のような品質面での課題が挙げられる。

- 要求定義と妥当性確認：機能および品質に関する要求は、モデル記法や自然言語記述を含む従来記法で表記可能か。仮に表記できたとして、実装された要求の妥当性をリリース前に確認することは



■ 図-1  
AI と従来ソフトウェアとの比較

特集  
Special Feature

可能か。

- 設計と検証：機械学習モデルの入力と出力に関する機能的な仕様が書けるか。仕様に基づく検証（レビュー，テスト）は実施可能か。
- 品質保証：品質の観点，メトリクス（測定量および測定方法），評価基準などは，従来型 ICT システムのものが適用可能か。

これら AI 応用システムの品質課題を踏まえ，その品質をどのような観点から評価し，担保するかについて議論が進んでいる。本特集で解説される国立研究開発法人産業技術総合研究所の「機械学習品質マネジメントガイドライン<sup>☆1</sup>（以降，AIQM）」<sup>2)</sup>や AI プロダクト品質保証コンソーシアムの「AI プロダクト品質保証ガイドライン<sup>☆2</sup>（以降，QA4AI）」<sup>3)</sup>はその代表的なものである。

本稿では，これらガイドラインに示された AI 応用

システムの評価観点を，ISO/IEC 25000 SQuaRE<sup>☆3</sup> シリーズ（以降 SQuaRE）で規定している従来型の ICT システムおよびデータの品質モデルに対応付けて考察し，その特徴的なポイントを浮き彫りにすることを試みる。最初に，SQuaRE を概説する。次に，上記ガイドラインの品質評価観点と SQuaRE の品質モデルとのマッピングを行い，対応関係や差異について考察する。最後に，AI 応用システムの品質に関する国際標準化の動向を紹介する。

## SQuaRE の概要

図-2 に示す通り，SQuaRE は，6 つの部門（Division）から構成される<sup>4)</sup>。

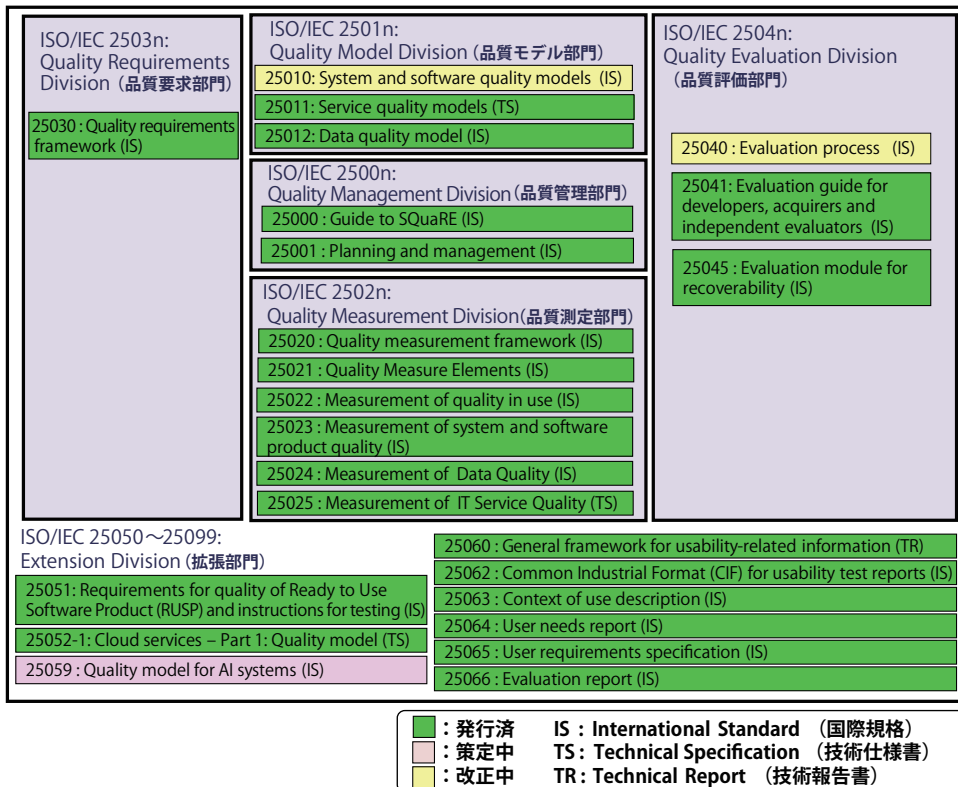
### ●品質管理部門

システムおよびソフトウェアのライフサイクルの中で，

☆1 Machine Learning Quality Management Guideline

☆2 Guidelines for Quality Assurance of AI-based Products and Services

☆3 Systems and software Quality Requirements and Evaluation



■ 図-2  
SQuaRE の体系



## 特集 Special Feature

SQuaRE で規定する品質モデル、品質測定量、プロセスをどの局面でどのように活用して品質を管理するかをガイドする。また、シリーズ全体を通して用いる用語や基本概念を規定する。現時点では、SQuaRE 適用の手引きと組織的な品質定量管理推進の手引きを提供している。

### ●品質モデル部門

評価対象や評価局面に応じて使い分けられる複数の品質モデルを規定する。品質モデルは、品質の概念を品質（副）特性と呼ばれる下位概念に展開し、各特性を定義する。現時点では、システムおよびソフトウェア製品、データ、IT サービスの3つの評価対象について、品質要求定義と品質評価の視点を与える品質モデルを提供している。システムおよびソフトウェア製品と IT サービスには、提供側の視点からのモデルおよび利用者側の視点からのモデルが含まれる。

### ●品質測定部門

品質特性を測定するために用いる品質測定量(quality measure)と測定方法(measurement method)を規定する。また、品質測定量を算出するのに用いる品質測定量要素(QME: Quality Measure Element)と測定方法を規定する。現時点では、品質測定の枠組み、各品質モデルに対応した品質測定量と測定方法、ならびにシステムおよびソフトウェア製品の品質測定量を構成するQMEと測定方法を提供している。

### ●品質要求部門

品質モデルと品質測定量を用いた品質要求事項の仕様化を規定する。現時点では、品質要求定義の枠組みを提供している。

### ●品質評価部門

品質モデルと品質測定量を用いた品質評価を規定する。現時点では、品質評価の基本要素事項、開発者、取得者、評価者の役割に応じた実践ガイドおよび評価ノウハウを蓄積するための枠組みを提供している。

### ●拡張部門

他の5つの部門で規定した品質要求定義および品

質評価に関する共通的、汎用的な事項を特定のコンテキストで実践するための、補完的な要求事項および推奨事項を規定する。現時点では、既製ソフトウェア製品(RUSP: Ready to Use Software Product)の品質認証、共通工業様式(CIF: Common Industry Format for usability)と呼ばれる使用性評価のための各種様式、およびクラウドサービスの品質モデルを提供しており、AIシステムの品質モデルを策定中である。

## SQuaRE 品質モデルとのマッピング

### マッピングの概要

ここでは、SQuaRE で規定している ICT システムの品質モデルと AIQM および QA4AI に示された AI 応用システムの品質特性および評価観点とのマッピングを示す。

ここで取り上げる SQuaRE 品質モデルは、ISO/IEC 25010 の利用時品質モデルと、ISO/IEC 25010 のシステムおよびソフトウェア製品品質モデル（以降、製品品質モデル）、および ISO/IEC 25012 のデータ品質モデルとする。これらのモデルと AIQM および QA4AI に示された AI 応用システムの品質特性および評価観点とのマッピングを図-3 に示す。図の中で、SQuaRE の利用時品質モデル、製品品質モデルは一部を除き品質特性が品質副特性に展開されている。それについては、対応付けの便宜のため“品質特性名/品質副特性名”形式で表した。

なお、SQuaRE 品質モデルでは、品質特性および品質副特性を定義する際に、「明示された状況下で使用するとき」といった形で、一定の利用条件の下で実装が求められる特質を規定している（要求を逸脱した状況での利用を議論する利用状況網羅性/柔軟性を除く）。他方、機械学習ソフトウェアでは、訓練データが仕様であるという見方はできるものの、利用条件は必ずしも明示的ではない。次節では、この前提条件を緩めた上で、関係性の考察を行う。

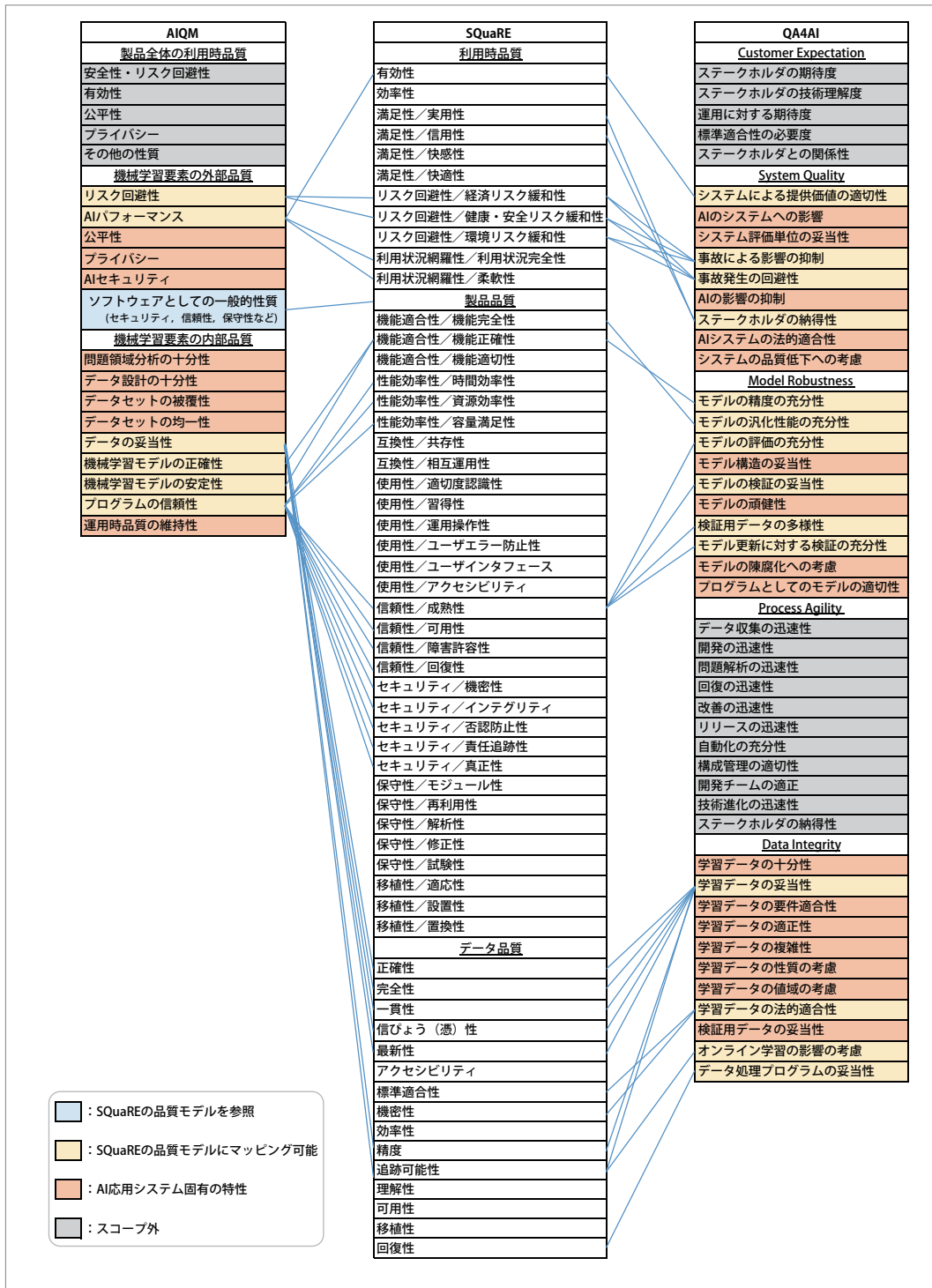
特集  
Special Feature

SQuaRE と AIQM

AIQM では、利用時品質は AI 応用システムが全体として利用時に満たすべきもの、外部品質は利用時品質の中で機械学習要素が満たすべきもの、内部品質

は外部品質を満たすために機械学習要素が固有に備えるべきものとしている。

SQuaRE との対比で考えると、AIQM の利用時品質と外部品質は SQuaRE の利用時品質に近く、AIQM の



■ 図-3  
AIガイドラインと  
SQuaRE 品質モデル  
とのマッピング



## 特集

## Special Feature

内部品質は SQuaRE の製品品質に近い。

以下、SQuaRE と AIQM とのマッピングについて考察する。ただし、AIQM の利用時品質は利用者が知覚する品質全般の意味合いで用いられており、定義が与えられているわけではなく、ここでの考察の対象から外す。

### < AIQM の外部品質 >

AIQM の外部品質特性のうち、まずは「リスク回避性」と「AI パフォーマンス」について考察する。

「リスク回避性」は、機械学習モデルの誤った判断により人的被害や経済的損失・機会損失などの悪影響を及ぼすリスクに関する品質特性としている。SQuaRE はリスク回避性を経済、健康・安全、環境の3つの要因に細分化している。一方、AIQM は環境リスクを明示していない。

「AI パフォーマンス」は、機械学習が提供する機能の有用性が重要視される分野への応用に着目した品質特性であり、SQuaRE の利用時品質モデルの利用状況網羅性、ならびに固有の利用状況のもとでの有効性に対応付けられる。

AI 応用システムでは、「リスク回避性」と「AI パフォーマンス」の双方を高いレベルで両立させることが理想であるが、現実的には、相反することが多い。そこで、応用分野によって「リスク回避性」と「AI パフォーマンス」の最適なバランスを要件として明確にすることが必要となる。

それ以外の外部品質特性「公平性」、「プライバシー」、「AI セキュリティ」に関しては、いずれも SQuaRE の利用時品質モデルのリスク回避性に概念的に関連するが、陽に対応付く特性はない。AI 応用システムでは、SQuaRE の「リスク回避性」の範囲を広げる必要がある。

### < AIQM の内部品質特性 >

データのモデリングに関連する「問題領域分析の十分性」と「データ設計の十分性」、データの分布に関連する「データセットの被覆性」、「データセットの均一性」は、SQuaRE の製品品質モデルにもデータ品質

モデルにも対応付く特性がない。これらの内部品質は、データモデリングで対応すべき状況の組合せを洗い出し、データ分布を考慮して、いずれのケースでも正しい推論を行うに足るデータを与えるとともに、全体としての推論性能を確保するためにデータセットの均一性を図ることであり、機械学習に固有である。

「データの妥当性」は、SQuaRE のデータ品質モデルの中から AI に関連するものを抽出し、機械学習の性質を考慮して解釈を見直したもので、SQuaRE のデータ品質モデルの正確性、完全性、一貫性、信ぴょう（憑）性、最新性、精度、追跡可能性と対応付けられる。

「機械学習モデルの正確性」、「機械学習モデルの安定性」は、SQuaRE の製品品質モデルの機能正確性と対応付けられる。

ただし、SQuaRE の機能正確性は入力データに対して正確な出力値が提供されることを扱うが、機械学習モデルでは機能正確性が十分かどうかを確認するのが簡単ではない。機械学習モデルの開発時には学習用の入力データに対して正確な出力値を提供したとしても、運用時に環境から与えられる入力データに対しては正確な値が必ずしも出力されない場合がある。AIQM では、正確性と安定性の両立が重要であるとし、安定性を向上するために適用した技術とその評価結果を記録することを機械学習モデルの品質管理における要求事項としている。

「プログラムの信頼性」は、SQuaRE の製品品質モデルの信頼性の副特性と対応付けられる。AIQM の「プログラムの信頼性」には、SQuaRE で規定した信頼性のほかに、リソースや時間の制約の充足、セキュリティに関する要求が含まれる。それらは、SQuaRE の製品品質モデルの性能効率性およびセキュリティの副特性と対応付けられる。

「運用時品質の維持性」は、名称の上では SQuaRE の製品品質モデルの保守性に近いようにも思えるが、訓練済み学習モデルの改版に伴う品質劣化の抑止などを扱う、AI 固有の特性である。

## SQuaRE と QA4AI

QA4AI では、AI プロダクトの品質保証において考慮すべき5つの軸として、「Customer Expectation」、「System Quality」、「Model Robustness」、「Process Agility」、「Data Integrity」を設定し、それぞれの分類軸における評価観点をチェックリスト形式で示している。

「System Quality」は、AI プロダクト全体の品質確保と価値提供に関するものであり、SQuaRE の利用時品質に近い。

「Model Robustness」は、モデルの精度、頑健性、デグレードなどに関するものである。システムおよびソフトウェア製品の品質特性/副特性を規定する SQuaRE の製品品質モデルと比べると、対象が AI 固有で限定的である。

「Data Integrity」は、学習用データの質と量の十分性や学習用データと検証用データの独立性などに関するものである。ICT システムで処理されるデータの品質特性を規定する SQuaRE のデータ品質モデルと比べると、対象が機械学習固有で限定的である。

「Customer Expectation」は、ステークホルダーニーズに関するものである。SQuaRE では、ステークホルダーニーズは利用時品質、製品品質、データ品質の要求事項を定義するための入力と位置付け、品質モデルとしては取り扱っていない。

「Process Agility」は、開発プロセスの品質に関するものである。SQuaRE では、プロセス品質<sup>☆4</sup>は製品品質への影響要因と位置付け、品質モデルとしては取り扱っていない。

本稿の冒頭に示したように、発注者のニーズに基づく仕様から演繹的にプログラミングを行う従来のソフトウェアと異なり、帰納的なプログラミングを行うことが機械学習の特徴である。このことから、機械学習を用いた AI 応用システムでは、前もって担保できる品質を予測し合意することが難しい。また、運用中に環

境が変化することによって精度が劣化することも起こりやすい。したがって、機械学習を用いた AI 応用システムでは、機械学習特有の性質を認識した上で発注者と合意すること、運用状況のフィードバックに基づき継続的な改善を行うことが重要となる。「Customer Expectation」は主に前者、「Process Agility」は主に後者に対応した評価観点を含む分類軸と考えられる。

以下、SQuaRE と QA4AI の「System Quality」、「Model Robustness」、「Data Integrity」とのマッピングについて考察する。

### < QA4AI の System Quality >

「システムによる提供価値の適切性」は、SQuaRE の利用時品質モデルの有効性に対応付けられる。ただし、価値を広義に捉えると利用時品質全体と対応付けることもできる。

「事故による影響の抑制」、「事故発生の回避性」は、いずれも SQuaRE の利用時品質モデルのリスク回避性の副特性と対応付けられる。

保証性や説明可能性を含む「ステークホルダーの納得性」は、SQuaRE の利用時品質モデルの実用性、信用性と対応付けられる。

その他は、AI 固有の評価観点である。「システム評価単位の妥当性」は、SQuaRE でもシステムアーキテクチャに基づく構成要素の評価に言及しているが、品質特性/副特性としては取り上げられていない。

### < QA4AI の Model Robustness >

「モデルの精度の充分性」は、処理や出力の正確さを規定する SQuaRE の製品品質モデルの機能正確性と対応付けられる。

「モデルの汎化性能の充分性」は、必要な作業や利用目的をどの程度網羅するかを規定する SQuaRE の製品品質モデルの機能完全性と対応付けられる。

評価・検証に関連する「モデルの評価の充分性」、「モデルの検証の妥当性」、「検証用データの多様性」、「モデル更新に対する検証の充分性」は、システムが十分に検証されて欠陥が除去されてどの程度長く正常に動作するかを規定する SQuaRE の製品品質モ

☆4 Process Agility の評価観点は、ISO/IEC 33000 シリーズ中のプロセスアセスメントモデルとの対応付けができるかもしれない。



## 特集

## Special Feature

デルの成熟性に対応付けられる。

その他は、AI 固有の評価観点である。

### < QA4AI の Data Integrity >

「学習データの妥当性」は、SQuaRE のデータ品質モデルの正確性、完全性、一貫性、信ぴょう（憑）性、最新性、精度、追跡可能性と対応付けられる。

「学習データの法的適合性」は、SQuaRE のデータ品質モデルの標準適合性、機密性と対応付けられる。ただし、SQuaRE では法的側面には言及しておらず、強制力は QA4AI の方が強い。

追加、置き換え、削除されるデータの監視、制御、検証を扱う「オンライン学習の影響」は、データ変更に対する監査証跡の提供を規定している SQuaRE のデータ品質モデルの追跡可能性と対応付けられる。

「データ処理プログラムの妥当性」は、故障時などにも一定水準の操作および品質を維持することを規定している SQuaRE のデータ品質モデルの回復性と対応付けられる。

その他は、AI 固有の評価観点である。「学習データの要件適合性」や「学習データの適正性」で言及されているデータのバイアスや公平性などに関する評価観点は、機械学習モデルの振舞いが学習データに大きく依存することを鑑みると、AI 品質の重要なポイントになる。

SQuaRE では、特定の利用状況でシステム、ソフトウェア製品、IT サービスを用いたときに、それらが利用者やステークホルダに与える成果や影響によって品質を評価する観点を利用時品質として整理している。現在、SQuaRE のデータ品質モデルでは、データの利用時品質については陽に言及していないが、AI 応用システムでは、学習データが利用時品質に与える影響が強くなると考えられる。ステークホルダの要求事項の充足やステークホルダと社会への影響を評価する「学習データの要件適合性」、「学習データの適正性」などは、データの利用時品質の枠組みの中で扱うのが適切かもしれない。

また、学習データの複雑性は、SQuaRE のデータ

品質モデルの理解性と通じるところがあるが、QA4AI では推論機能に対するデータの複雑さや過度な単純化を議論しているのに対し、SQuaRE ではデータの利用者にとっての理解性を規定しており、主体が異なっている。

## AI 品質の国際標準化動向

AI に関する国際標準化は、2017 年に設置が決議された ISO/IEC JTC 1/SC 42 : Artificial Intelligence<sup>5)</sup>で審議が進められている。主な WG の構成は、次の通りである。

- WG 1 : Foundational standards
- WG 2 : Data
- WG 3 : Trustworthiness
- WG 4 : Use cases and applications
- WG 5 : Computational approaches and computational characteristics of AI systems

これまでに、13 の国際規格 (IS) や技術報告書 (TR) が発行され、25 のプロジェクトが進行中である。

この中で、SQuaRE と関連の深いテーマを扱っているのは、WG 2 と WG 3 である。WG 2 のいくつかのプロジェクトは、SQuaRE で規定しているデータ品質モデル (ISO/IEC 25012) およびデータ品質測定量 (ISO/IEC 25024) と関連している。WG 3 のいくつかのプロジェクトは製品品質モデルおよび利用時品質モデル (ISO/IEC 25010) および品質評価のガイドライン (ISO/IEC 25040, 25041) と関連している。それらのプロジェクトのステータス<sup>☆5</sup>とタイトルを表-1 に示す。

WG 2 は、機械学習に使われるデータの品質を扱っており、ISO/IEC 5259-2 はデータの品質の測定量を規定しようとしている。ISO/IEC 5259 シリーズには

<sup>☆5</sup> 国際規格の制定は次の段階を経て発行に至る。PWI (Preliminary Work Item) =>NP (New Work Item Proposal) =>AWI (Approved Work Item) =>WD (Working Draft) =>CD (Committee Draft) =>DIS (Draft International Standard) =>FDIS (Final Draft International Standard) =>IS (International Standard)

## 特集 Special Feature

データ品質モデルを含まないが、ISO/IEC 5259-2 に規定される品質測定量は、ISO/IEC 25012 で規定している品質特性と関連付けて規定されることが望まれる。

WG 3 は、Trustworthiness の一部として、AI システム品質を取り扱っている。本稿のテーマと特に関連深いのは、ISO/IEC 25059 : Quality model for AI systems<sup>6)</sup>である。この規格に関しては、プロジェクトの開始に先立って、SC 7/WG 6 (SQuaRE を担当) コンビナーを務める筆者 (込山) と同プロジェクトリーダーの Adam Leon Smith 氏との間で協議し、この規格を SQuaRE シリーズの 1 規格とすること、

ISO/IEC 25010 の適合性 (Conformance) の要求事項を満たすことを合意した。この適合性は、品質モデルを国際規格に準拠した形でカスタマイズする際、変更個所とともに、その変更が必要な理由を明示するというものである。

この規格は、現在 DIS 段階まできており、ISO ポータルから入手可能である。ISO/IEC 25010 の製品品質モデルおよび利用時品質モデルに追加または修正された品質特性/副特性およびそのポイントを表-2 に示す。

■表-1 SQuaRE との結びつきが強い SC 42 のプロジェクト

WG	ステータス	文書番号	タイトル
2	AWI	5259-1	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 1 : Overview, terminology, and examples
2	AWI	5259-2	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 2 : Data quality measures
2	AWI	5259-3	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 3 : Data quality management requirements and guidelines
2	AWI	5259-4	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 4 : Data quality process framework
2	AWI	5259-5	Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 5 : Data quality governance
3	AWI	TS 5471	Artificial intelligence — Quality evaluation guidelines for AI systems
3	DIS	25059	Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

■表-2 AI システムの品質モデルで追加・修正された特性

モデル区分	品質特性/品質副特性	追加	修正	追加・修正のポイント
製品品質モデル	Functionality/ Functional adaptability	X		新たな学習データ、運用時の入力データ、システムの事前の動作結果などから学習し、動的に変化する環境に適応できること
	Functionality/ Functional correctness		X	AI システム内の確率的な機械学習方式は出力にある率で誤りが生じること
	Usability/ User controllability	X		AI システムの不測の動作の際に人や外部エージェントが介入できること
	Usability/Transparency	X		AI システムが利害関係者の利用目的や要件に合致するかを判断できる情報を提供すること
	Reliability/Robustness	X		不測のデータ入力、バイアスのかかったデータ入力、外部からの干渉などがあっても AI システムが一定水準の正確性を維持すること
	Security/Intervenability	X		AI システムが危害や危機を生じることを予防するためにその機能に介入できること
利用時品質モデル	Satisfaction/ Transparency	X		Usability/Transparency と同じ
	Freedom from Risk/ Societal and ethical risk mitigation	X		アカウンタビリティ、公平性、順法など、社会的および倫理的リスクを緩和すること



## 今後の課題

AI 応用システムの利用が進むのに伴って、AI の潜在的な問題が及ぼす、社会、倫理などへの影響が懸念されるようになり、国際機関、地域、国などで、規則や原則を制定、整備する動きが出てきている。下記はその例である。

- EU : 「AI 規則案 (EU AI Act)」
- OECD : 「人工知能に関する OECD 原則」
- 日本 : 「人間中心の AI 社会原則」

どの品質特性がこれらの規則や原則の充足につながるか、規則や原則への適合性を品質モデルにどう組み入れるかなど、これらの規則や原則と AI システムの品質モデルとの関連には関心が及ぶところであるが、その考察は別の機会としたい。

国際標準化に関しては、AI 品質に関する国際規格が発行された際には、それらを利用して AI 応用システムの質の向上を促進するとともに、その適用結果を国際標準にフィードバックして、より実務に役立つ国際規格に昇華させていく予定である。このような取り組みを経て、AI 品質に関する技術的な基盤が国際規格群として整備され、それらを活用して、必要な規則や原則が制定され、質の高い安全・安心な AI 応用シ

ステムが構築・活用され、世の中が発展していくことが望まれる。

### 参考文献

- 1) 丸山 宏：機械学習工学に向けて、日本ソフトウェア科学界第 34 回大会講演論文集 (2017)。
- 2) 国立研究開発法人産業技術総合研究所：機械学習品質マネジメントガイドライン第 3 版 (2022)。
- 3) AI プロダクト品質保証コンソーシアム：AI プロダクト品質保証ガイドライン (2022. 07 版)。
- 4) 込山俊博，東 基衛：システムおよびソフトウェア品質の国際的な基準の確立—日本主導の国際標準化への取組み—，デジタルプラクティス，Vol.10, No.1 (Jan. 2019)。
- 5) ISO/IEC JTC 1/SC 42 - Artificial Intelligence : <https://www.iso.org/committee/6794475.html>
- 6) ISO/IEC DIS 25059: Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) — Quality Model for AI Systems (2022)。

(2022 年 7 月 29 日受付)

#### ■込山俊博 (正会員) t-komiyama@nec.com

NEC ソフトウェア&システムエンジニアリング統括部 ICT 標準化エキスパート。1985 年慶應義塾大学 工学部 数理科学科卒業後、NEC 入社。ISO/IEC JTC 1 SC 7/WG 6 コンビナー。米 CMMI Institute 認定 CMMI リードアプレイザ/インストラクタ。独 intacs 認定 Automotive SPICE プリンシパルアセッサ。

#### ■向山 輝 amukai@nec.com

NEC ソフトウェア&システムエンジニアリング統括部。1990 年上智大学大学院理工学研究科博士前期課程修了後、NEC 入社。LSI 設計自動化、形式手法などの研究開発に携わり、現在は、機械学習システムの開発・運用プロセスの策定に従事。

