機械学習を用いたシステムの品質管理におけるベストプラクティスについて

丸山 宏

機械学習を用いたシステムの品質管理についてはいくつかのガイドラインが存在するが，それらをどのように現実のプロジェクトに適用していくかについてはまだ多くの試行錯誤が必要と考えられる．本稿では，実際にビジネスに用いられている 3 つのプロジェクトについて，その品質マネジメントの実践をレビューし，それらの共通点と相違点を明らかにする．その上で，機械学習を用いたシステムの品質管理には，ドメインに応じたハザード列挙の方法論が必要であることを議論し，そのためのベストプラクティス集の編纂を提案する．

1 はじめに

機械学習を使ったシステムは，今までのソフトウェアの作り方と大きく違うために，その品質管理が難しい [6]．このため，いくつかの品質ガイドラインが制定されてきた [5] [1]．それらをどのように現実のプロジェクトに適用していくかについてはまだ多くの試行錯誤が必要と考えられる．本稿では，実際にビジネスに用いられている 3 つのプロジェクトについて，その品質マネジメントの実践をレビューし，それらの共通点と相違点を明らかにする．その上で，機械学習を用いたシステムの品質管理には，ドメインに応じたハザード列挙の方法論が必要であることを議論し，そのためのベストプラクティス集の編纂を提案する．

本稿ではまず 2 章で，本稿で考える品質とは何かを議論する．3 章では既存の品質ガイドラインをレビューする．その上で，4 章で実際にビジネスで運用されている 3 つのシステムを取り上げ，それぞれどのような考え方で品質管理がされているかを明らかにする．これらのケーススタディで明らかになるのは，ドメインごとにハザード列挙の考え方が大きく異なることである．この仮説に基づき 5 章ではドメインに基づくベストプラクティス集の編纂が必要であることを議論する．

2 品質とは

製造業における製品品質が語られていた初期には，品質とは製品にばらつきのないこと，と大まかに理解されていたが，品質の概念がサービスなど無形物にも拡大されるようになってからは，品質とは「対象（製品・サービス）が持っている特性が，顧客の期待を満たしている度合い」と理解されることが多くなった．品質特性としては，機能・性能・信頼性などの他に，セキュリティ・プライバシー・公平性・安全性など様々なものが考えられ始めている．最近では，特に生成 AI の文脈において**アラインメント**という言葉が聞かれるようになったが，これも広くは品質特性の一部と見ることもできる．

このように，品質とは対象の特性だけで決まるのではなく，対象に対するステークホルダの期待との相対関係で決まることに注意が必要である．このことは，いわゆる「AI」という言葉で過大な期待や警戒感を持つステークホルダがいる，機械学習系のシステムにおいては重要であると言える．

しかしながら，顧客の期待を事前に明確にするのは極めて困難であることは，ソフトウェア工学の長い歴

On Best Practice of Quality Management for ML-Based Systems

Hiroshi Maruyama, 花王/東大/Preferred Networks, Kao Corp./U. Tokyo/Preferred Networks.

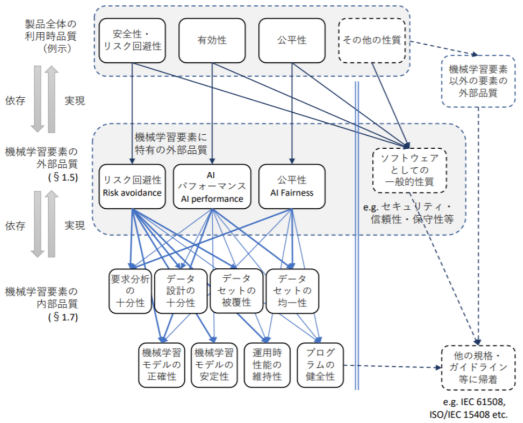

図 1　産総研の機械学習品質マネジメントガイドライン
機械学習コンポーネント特有の品質属性として、3 つの外部品質と 8 つの内部品質を定義している

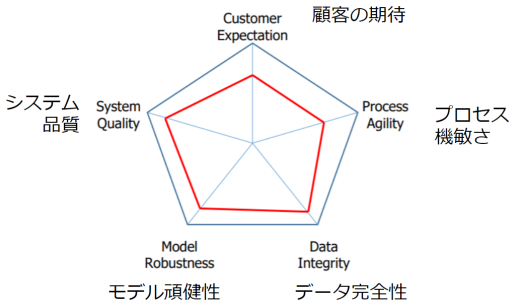

図 2　AI 品質マネジメントガイドライン
機械学習コンポーネントを含むシステム全体の品質を 5 つの軸で評価する

史が示している (アジャイル開発は，顧客の要求が事前にわからないことを前提に作られた手法だといえる)．このため，品質管理においては，「起きてはまずいこと」(これを**ハザード**と呼ぶ) をできるだけ漏れなく列挙することが大切となる．

可能なハザードを全部列挙することは到底できないが，それでも列挙されたハザードがどの程度起きやすいか，その結果がどれほど致命的か，によって**リスク**を評価することができ，そのリスクの程度によってどのようなリスク軽減策を取れるか，を決めるのが品質管理の要諦となる．

3　機械学習を用いたシステムの品質ガイドライン

我が国では比較的早くから深層学習など機械学習を使ったシステムの品質管理に議論が行われていて，世界に先駆けて産総研の機械学習品質マネジメントガイドライン [5] や，AI プロダクト品質保証コンソーシアムによる AI 品質マネジメントガイドライン [1] が作られ，公開されてきた．

図 1 に示す機械学習品質マネジメントガイドラインは，機械学習コンポーネントに対して既存の品質特性に加えて，リスク回避性・AI パフォーマンス・公平性という 3 つの追加の品質特性を評価するように，とガイドしている．機械学習コンポーネントのその他の品質特性，及びシステムの他の部分の品質特性については，既存のソフトウェア工学における一般的な品質特性，例えばセキュリティや信頼性で評価すべき，という考え方である．

一方，AI 品質マネジメントガイドラインでは，機械学習コンポーネントだけではなく，機械学習コンポーネントを含んだシステム全体の品質を，システム品質・モデル頑健性・データ完全性・プロセスの機敏さ・顧客期待の 5 軸で評価することを求めている．

この 2 つのガイドラインだけを見ても，それぞれのフォーカスが大きく異なる．また，機械学習品質マネジメントガイドラインが定義する公平性という品質特性は，人間が介在しない機械装置に使われる場合には適用外になるなど，ドメインにおける適合性が大きく異なる点も使いづらいところである．さらに，機械学習の技術そのものが急速に変化していて，どちらのガイドラインも頻繁に改訂を受けていることも，静的な文書としてのガイドラインを品質管理の主要な拠り所とすることの難しさの原因になっている．

4　機械学習を用いたシステムの品質管理の事例

本章では，実際のビジネスに使われている機械学習を用いたシステムを 3 点取り上げ，公表されている情報からそれらの品質管理の考え方を抜き出す．

4.1　システム A：機械学習コンポーネントが組み

込まれた工作機械

第 1 のシステム（これを**システム A** と呼ぶ）は，工場で使われる工作機械であり，その 1 部に機械学習コンポーネントが組み込まれている [2]．システム A における品質管理プロセスを図 3 に示す。

品質管理を行う主体はこの工作機械を製造し販売するメーカーである．このため，この論文での品質管理は，製品の企画・設計・製造から出荷までのプロセスが対象範囲であり，装置の購入・設置・運用・保守については範囲外となっている．

ハザードの初期の数え上げとしては，品質管理の対象システムが機械装置であることから，宇宙航空分野など高信頼機械システムで使われる手法である FMEA(Failure Mode - Effect Analysis) を用いている．FMEA では，対象システムの各要素の機能を洗い出し，「この機能が失われたら？」のようなガイドワードでハザードを見つけていく．この際，3 章で述べた QA4AI による AI 品質管理ガイドラインを参照してハザードのガイドワードを拡張した．

さらに，故障モードから導かれたリスクに対して，リスクの優先順位をつけた後に，逆向きにそのリスクにつながる他の要因も探索することで，よりカバレージを広げている．

4.2 システム B：ライフケア生成モデルの API サービス

人体に関わる様々な計測値の確率同時分布を API として提供するサービスであり，ヘルスケア・ライフケア分野における汎用的な統計モデルとして利用されることを狙ったものである [4]．

システム B における品質管理プロセスを図 4 に示す。このプロセスは、構築から運用までシステムのライフサイクル全般にわたる品質管理を扱っていて，そのためモデルの継続的な改善やその評価指標，さらにはプロセス全体のガバナンスなど組織的な観点までをカバーしている．

この論文では，ハザードの数え上げを，データ被験者，データセット提供者，アプリケーション事業者，エンドユーザーという，ステークホルダ単位で行っていることに特徴がある．また，ヘルスケアに関連することから，薬機法などコンプライアンスに強い重点が置かれていること，それに関連して顧客期待のコントロールとデータ・アルゴリズムの透明性を明確にしている点に特徴がある．

4.3 システム C：化学プラントにおける自動運転

システム C は石油精製プラントにおける世界最初の自動運転システムであり，原材料流量変更に伴う運転状態の変動を自動的に制御する [3]．制御の誤りは，巨大なプラントの停止や事故につながる可能性があるために，品質が要求される．このため，化学プラントではもともと HAZOP によるリスク管理が一般化している．HAZOP においては，FMEA とは異なり「この値が既定値より大きくなったら」のようなガイドワードを利用する．

この論文では，HAZOP に加えて，機械学習コンポーネントに対してプラント保安分野 AI 信頼性評価ガイドライン [7] を用いてシステム評価を実施した．このガイドラインは，3 章で述べた産総研の機械学習品質マネジメントガイドラインをベースに，プラント保安分野への拡張を行ったものであり，機械学習コンポーネントのシステムへの組み込み方をいくつか類型化して議論している．その類型の 1 つとして，既存のルールベースのシステムによって，機械学習コンポーネントの望ましくない出力をフィルタする，というものがあり，このプラント自動運転システムはその類型に相当する (図 5)．すなわち，HAZOP に基づく「ガードレール」があることで，機械学習コンポーネントに起因するハザードを取り除いている．

5 ベストプラクティスの重要性

以上見てきたように，汎用のガイドラインだけで機械学習を使ったシステムの品質を効果的に維持するのは困難である．それぞれのドメインには，そのドメインでの経験・知識に基づいた品質管理の手法があり，その中にガイドラインから得られた知見をうまく埋め込むことで，効果的な品質管理ができると考えられる．

各ドメインで異なる品質特性が重視されること，基盤モデルの利用など機械学習の技術そのものが急速に

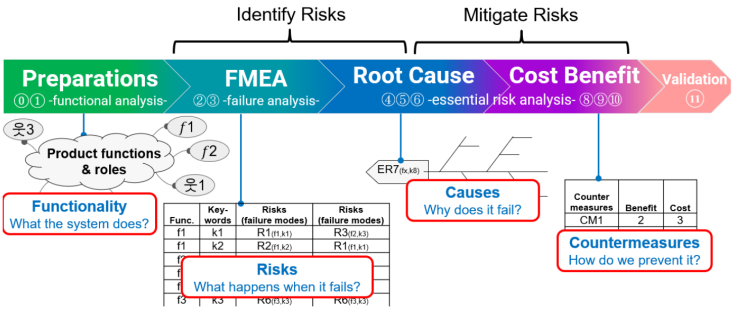

図 3　システム A における設計時の品質管理プロセス

FMEA でハザードを数え挙げた上で、その根本原因を洗い出し、リスクの大きさと緩和策のコストに基づいて対策を決定する

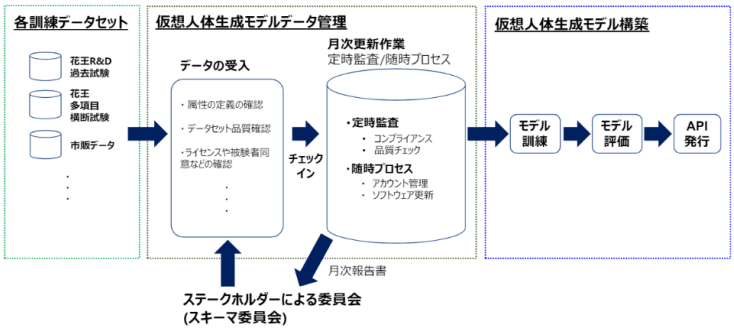

図 4　システム B における品質管理プロセス

訓練データの整備・モデル訓練・評価というプロセスと、それを管理するガバナンス体制を整備している

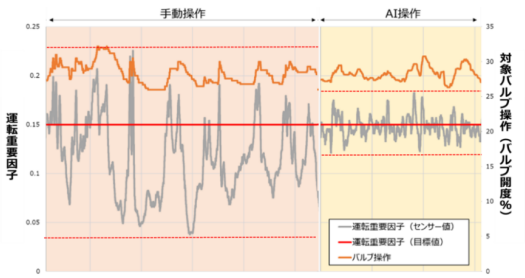

図 5　システム C における安全管理

バルブ開度が HAZOP に基づく管理基準範囲 (赤の点線) に収まるように監視している

変化しつつあることを考えると，汎用性のあるガイドラインを整備することよりも，それぞれのドメイン，それぞれの文脈で柔軟に品質管理を設計した，ベストプラクティス集を整備することのほうが効果的ではないかと考える．

新たな機械学習を組み込んだシステムの設計者は，このベストプラクティス集を見て，自分が担当するシステムに似た品質コンサーンを持つ事例から，役に立ちそうな手法を採用すればよい．

品質管理で重要なのは，いかに致命的なハザードを見逃さないか，である．ハザードを漏れなく列挙する形式的な手法がない現段階においては，これはプロジェクトの利害関係者の想像力によるしかない．そのための方法論が，このベストプラクティス集に広く紹介されていれば，より効果的な品質管理ができるのではないか．

謝辞

本稿で紹介した各プロジェクトの関係者に謝意を表します．

参 考 文 献

[1] 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 2024.04 版, 2024.

[2] Dominguez, G. A., Kawaai, K., and Maruyama, H.: Quality Assurance for ML Devices A Risk-Based Approach, *2023 30th Asia-Pacific Software Engineering Conference (APSEC)*, IEEE, 2023, pp. 524–531.

[3] 平井太一朗ほか: AI によるプラント自動運転の実証と信頼性評価について, 第 *54* 回安全工学研究発表回講演予稿集, Tokyo, 2021.

[4] 尾藤宏達ほか: 仮想人体生成モデルにおける品質管理, 日本ソフトウェア科学会第 *40* 回大会講演論文集, Tokyo, 2023.

[5] 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 4 版, 2024.

[6] 石川ほか (編): 機械学習工学, 講談社, Tokyo, 2022.

[7] 石油コンビナート等災害防止 3 省連絡会議: プラント保安分野 AI 信頼性評価ガイドライン 第 2 版, 2021.