船舶へのAIシステム活用における，開発過程の品質管理について

宮島　秀規*

1. はじめに

近年海事業界においても，AIシステムの開発が進んでおり，例えば，画像認識技術を利用した状況認識支援システム[1) 2) 3)]の開発や，避航操船，自動離着桟を行う自動操船AIの研究・実証実験[4) 5) 6)]などが実施されている。今はまだ船員の補助としての使われ方や，研究段階のAIがメインであるものの，将来的には，船舶に搭載されたAIシステムが，安全に直結する形で活用されるようになることが考えられる。

このようなAIには，データからパターンやルール等を学習する機械学習の技術が用いられている。また，昨今特に注目を集めている深層学習は，複雑なモデルを膨大なデータで学習させることで高い性能を発揮している。しかしながら，中身は実質ブラックボックスとなっており，判断根拠を人間が理解することは難しい。また，確率的な推論を行うAIを実環境に導入する場合，開発者の意図していない結果を出力する可能性について考慮する必要がある。

AIシステムについて，品質保証の観点から考えると，データを基に学習することでふるまいが帰納的に決定される仕組みとなっている昨今のAIは，内部設計から導き出される出力値を明示的に説明することが難しい。そのため，期待通りの性能が発揮できるかどうか，評価するのは容易ではない。また，開発後も周囲の状況の変化に追従できるように継続的な学習が行われる場合が多い。そのため，設計から運用までを含めたライフサイクルを考慮した開発が求められる。そこで，まずは開発過程で品質を確保するために必要な活動が行われているか，という点を押さえ，そのうえで具体的な評価方法を検討していく必要がある。

このようなAI特有の事情を踏まえ，本会では，適切な品質管理の下開発されたAIシステムが海事業界で利用されることを企図して，AIシステムの開発者が開発時に考慮すべき事項を『船舶へのAIシステム活用に向けたテクニカルガイド -AIシステムの開発過程における品質管理-』として取りまとめる予定である。そこで本稿では，現在開発中のガイドの内容として，AIの概要や開発と運用を一体化したアプローチについて説明する。また，今後海事業界におけるAIシステムの開発プロセスを整理する際に参考になる考え方として，『機械学習品質マネジメントガイドライン』における開発プロセスの整理を紹介する。

2. AIについて

2.1　AIとは

AIという言葉は1956年のダートマス会議において提案され，その後様々な研究が行われてきた。AIの定義については様々な議論がなされているものの，確立した定義はなく，“人間の思考プロセスと同じような形で動作するプログラム，あるいは人間が知的と感じる情報処理・技術”というような広い概念で理解されている[7)]。

AIには探索や推論のアルゴリズム，エキスパートシステムなど様々な技術が含まれているものの，近年のブームの中心は機械学習である。機械学習はデータを分析する方法の一つであり，人間の学習に相当する仕組みをコンピュータ等で実現している。計算方法（アルゴリズム）に基づき，入力されたデータ（学習用データ）からコンピュータがパターンやルールを発見し，そのパターンやルールを新たなデータに当てはめることで，その新たなデータに関する識別や予測等を行うことができる。

また，機械学習の手法の一つに深層学習があり，近年のAIブームのきっかけとなった手法として特に注目されている。深層学習では人間の脳内の神経細胞の仕組みを再現したニューラルネットワークを用いており，そのニューラルネットワークを多層構造にしていることが特徴である。

*　開発本部　技術研究所

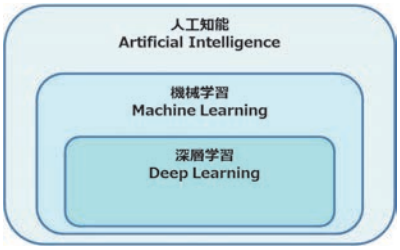

図1　AI・機械学習・深層学習の関係[7)]

2.2　深層学習

深層学習のニューラルネットワークは入力層，中間層，出力層で構成されており，入力されたデータからニューラルネットワーク内での計算を通じて特徴を抽出し，最終的な判断結果を出力する。ニューラルネットワークの計算を行うためのパラメータはデータを基に最適化を行っており，そのプロセスを学習と呼ぶ。

深層学習ではネットワークを多層化し，大量のデータで学習することで，様々な分野において従来手法より高精度な結果を達成している。ネットワークの構造にも，画像の扱いに特化したConvolutional Neural Networks (CNN)や，時系列データやテキストデータなどを扱う際に使用されるRecurrent Neural Networks (RNN)，与えられたデータセットから新しいデータを生成する生成AIなど様々なものがあり，用途によって使い分けられている。

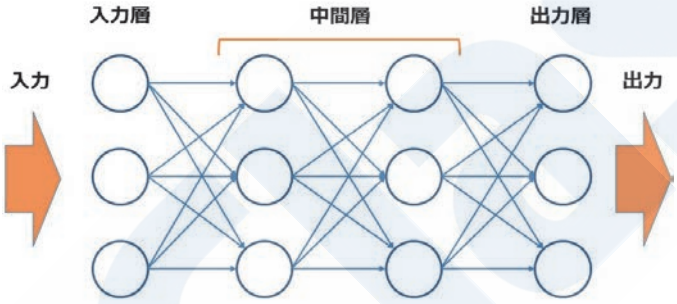

図2　深層学習の仕組み[8)]

2.3　AIの課題

機械学習を中心としたAIは精度の高い予測を行うことができるため，自然言語処理や音声処理，画像認識，コンテンツ生成など幅広い分野で応用されている。一方で，大量のデータを基に学習しており，また基本的には確率論で判断をしているため，以下のようなAI特有の課題について対応する必要がある。

2.3.1　データの品質と量

AIはデータを基に学習するため，期待通りの性能を引き出すにはデータの品質や量が重要になる。例えば，過去の規則性が当てはまらない場合や，判断基準に使われている要素がデータに含まれていない場合，データ量が少ない場合には，AIに十分な精度が期待できない可能性がある。また，学習データに偏りがあると，モデルが出力する推論結果に誤差が生じてしまう可能性もある。そのため，質的，量的に十分な量のデータを確保する必要がある。そこでAIシステムの開発時には，解決しようとしている問題をしっかりと分析し，目的や使用環境，条件などを明らかにしたうえで，必要となるデータの品質や量についての検討を行っていくことが重要となる。

2.3.2　計算資源やコスト

広範囲に及ぶデータの収集やアノテーションには膨大な時間を要する。また，ビッグデータを保存し，効率的に利用できるようにするためには大規模なストレージが必要となる。さらに，AIモデル，特に深層学習のモデルを学習させるためには大量の計算資源が必要であり，学習にかかる時間も考慮する必要がある。そのため，十分な開発体制や開発環境を整えることが求められる。

2.3.3　ドリフト[9)]

機械学習においては，時間経過によってモデルの予測性能が劣化するドリフトという現象が起きる。ドリフトには主に2種類あり，季節性やトレンドの変化のような，学習時のデータと運用時のデータの分布にずれが生じるものをデータドリフト，今までになかったものが登場した場合など，入力データと正解ラベルの関係性が学習時と比べて変化するものを概念ドリフトと呼ぶ。このようなドリフトに対処するためには，AIシステムを開発した後も，性能や運用環境の変化を監視し，定期的に再学習することが求められる。

2.3.4　確率的なふるまい

確率的な推論を行うAIを利用する場合は，間違った推論結果が出力される可能性があることを考慮する必要がある。また，生成AIにおいては，AIが真実に基づかない情報を生成するハルネーション[10)]という現象が発生する点にも注意が必要である。AIシステムを開発する際は，このようなAIの特徴についてステークホルダーに十分に理解してもらうための活動や，適切な目標，運用方法について検討することが重要になる。

2.3.5　ブラックボックス問題

機械学習のモデル，特に深層学習はパラメータ数が多く構造も非常に複雑であり，どのような根拠により判断を行ったかを人間が理解することは難しく，中身は実質ブラックボックスとなっている。このようなAIは高い精度での予測・認識を行うことがで

きるものの，「中身が説明できないものは使えない」という懸念がAIシステムへの不信につながり，導入を阻害する要因にもなりうる。そこで，ステークホルダーからの理解を得るための活動や，ブラックボックス問題を考慮した運用方法の検討，ブラックボックスではない説明可能なAIの使用の検討を行うことが重要になる。

2.3.6　評価の難しさ

データを基に学習するAIは，定義された仕様に対して内部設計や実装を明示的に関連付けることができず，内部設計や実装の良し悪しのレビューや品質の評価が難しい。また，技術の進歩が速く，活用方法も様々である。そのため，システム全体で達成したい目的やAIシステムに求める要素を明らかにし，対象とするAIシステムにあわせた評価方法を検討していくことが重要になる。

2.3.7　活用方法によるリスクの違い

AIは活用方法によってリスクが大きく異なり，翻訳や質問応答，音声認識のような日常生活で使用するAIシステムであれば，大きな問題は生じないと考えられる。一方で，自動運転や自動操船のような安全に直結するAIシステムの場合は，安全性や信頼性の面について慎重に検討する必要がある。

欧州委員会が2021年4月21日に発表したAI規則案ではリスクベースのアプローチが採用されており，4つのリスクレベルを設け，各々のリスクに応じた要件・規制が規定されている[11]。AIシステムを開発する際にはこのような考え方が参考になる。リスク評価を通じてAIシステムのリスクを把握し，適切な制御対策をとることで，安全性を担保していくことが重要である。

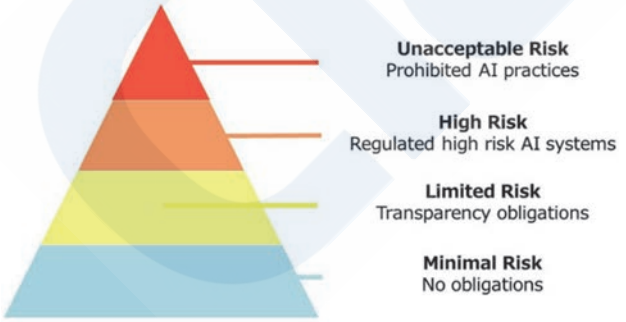

図3　EUのAI規制法案におけるリスクのピラミッド[12]

2.3.8　セキュリティ[13]

AIシステムを開発する際には，従来のITやシステムのセキュリティの観点に加えて，AI特有のセキュリティとして学習用データやAIモデルを狙った攻撃について考慮する必要がある。AIモデルへの攻撃には以下のようなものがあり，AIシステム開発時には，リスク評価の実施や，攻撃への対応策をとることが求められる。

・ポイズニング攻撃（Poisoning Attack）

学習用データやAIモデルに何らかの細工をして，AIモデルの開発者の意図しない推論結果を出力させる攻撃の一種。

・回避攻撃（Evasion Attack）

推論用データにノイズが加えられることでAIモデルの推論が誤って行われることがある。加えられたノイズが小さい場合，人間が元のデータとの差異を識別できないことがある。そうした小さなノイズが加えられたデータを敵対的サンプルと呼び，敵対的サンプルを利用してAIの誤認識を発生させる攻撃を回避攻撃と呼ぶ。

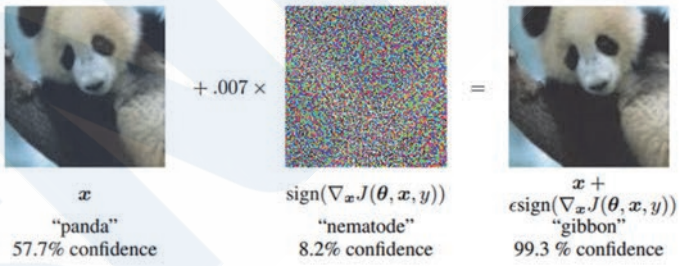

図4　テナガザルと誤認識されるパンダの敵対的サンプルの例[14]

実際にAIを活用したシステムを開発する場合は，開発プロセスにおける活動を通じて，このような課題に対応していく必要がある。

3. 開発と運用を一体化したアプローチ

一般的にAIシステムでは，期待通りの性能が得られるように，データを基にモデルの学習と評価を繰り返す形で開発が進められる。また，開発したAIシステムを実環境に導入した後も性能を監視し，ドリフトのような周囲環境の変化に起因するデータの傾向の変化への対応や，新たに蓄積したデータを用いた精度向上を図るために，継続的な評価や学習が行われる。そこで，AIシステムの開発においては，新たに取得したデータでの学習を繰り返すことで品質を維持・向上し，より期待する成果に近づけていく，開発と運用を一体化したアプローチ[15) 16)]が採用されることが多い。このような点を踏まえ，ガイドでは，AIシステムの開発初期から運用終了までのアプローチを，図5の通り整理する予定である。

実際の開発の流れは開発者や開発するシステムに

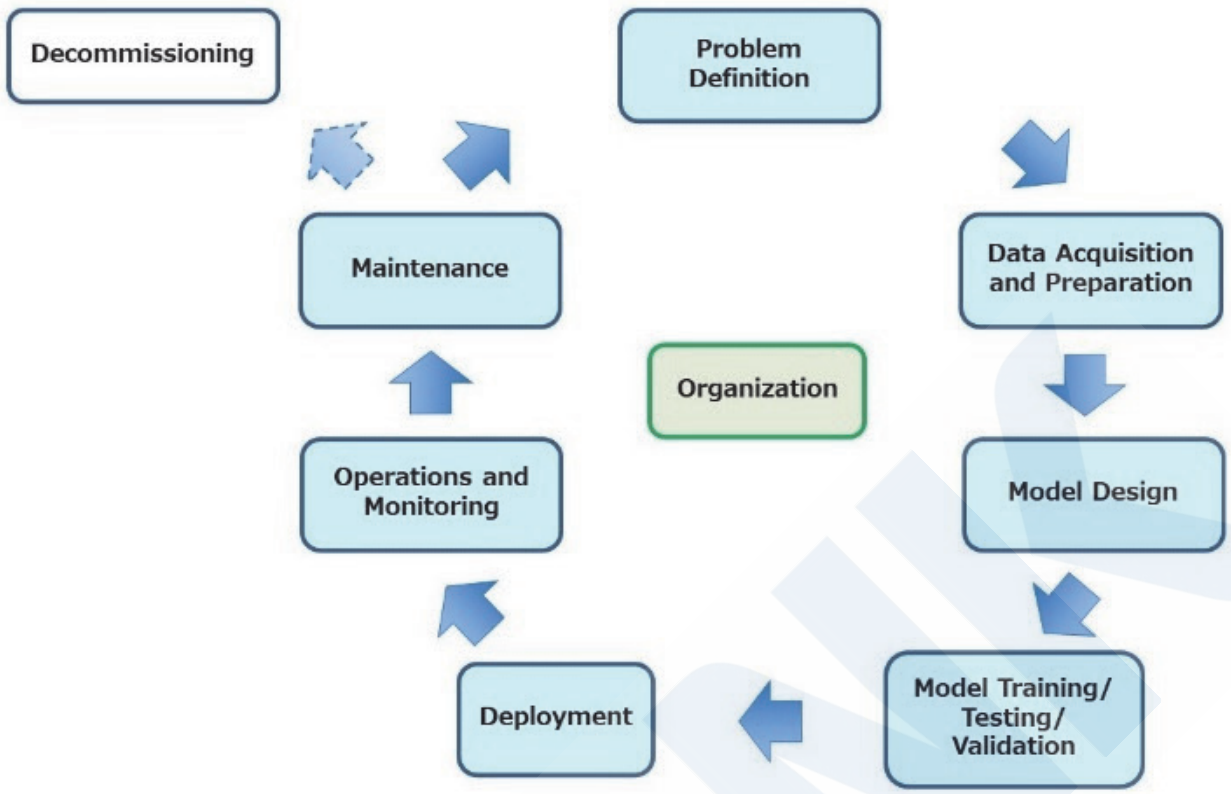

図5　開発と運用を一体化したアプローチ

よって異なるため，ここでは様々なAIシステムを意識した汎用的なアプローチを示している。AIシステムの開発においては，このようなアプローチの各フェーズを通じて，前述したAI特有の課題に取り組みつつ，品質を確保するために必要な活動を実施していくことが求められる。そこで，各フェーズで開発時に考慮すべき事項をガイドに記載することを検討している。

AIシステムの開発においては開発体制や開発環境も重要であり，アプローチ全体に関係する要素であることから，Organizationとして中心に配置している。

0) Organization
AIシステムの開発を成功に導くために必要な，開発，運用に携わる組織の能力に関する要素。

1) Problem Definition
AIシステムで解決しようとしている問題を明確にして，開発を進めるために必要な検討を行うフェーズ。

2) Data Acquisition and Preparation
データを収集し，AIモデルの学習に使用するための前処理を行うフェーズ。

3) Model Design
Problem Definitionで定めた目的や要件，利用可能なデータを踏まえて，開発の全体的なアプローチを定め，AIモデルを設計するフェーズ。

4) Model Training / Testing / Validation
AIモデルを学習させ，期待通りに動作していることを確認するフェーズ。

5) Deployment
実際の環境で動作するように，AIシステムを既存のシステムやプロセス，製品，サービスと統合するフェーズ。

6) Operations and Monitoring
AIシステムを実際の環境で運用し，期待通りに動作しているか継続的に監視するフェーズ。

7) Maintenance
必要に応じてメンテナンスやシステム，モデルの更新を実施するフェーズ。

8) Decommissioning
AIシステムの運用結果を評価し，AIシステムが不要になるか，もしくは，別の効果的なソリューションが開発された場合に，AIシステムの運用を終了するフェーズ。

4. 機械学習品質マネジメントガイドラインにおける開発プロセスの整理

前述の通り，開発と運用を一体化したアプローチ

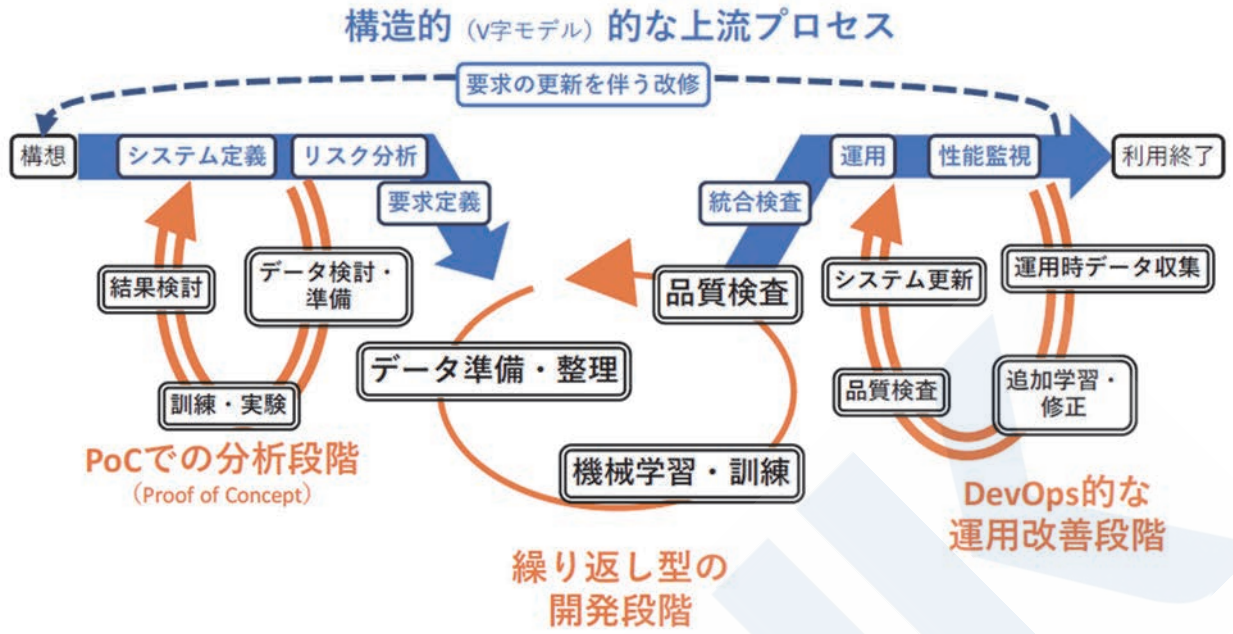

図6　機械学習品質マネジメントガイドラインにおける開発プロセスの整理[15)]

は対象を問わない汎用的なものとなっており，継続的に学習を繰り返し，品質を高めていくことを想定している。実際に船舶に搭載するAIシステムに適用する場合は，海事業界の開発プロセスに適した形に落とし込む必要がある。このような場合に参考になる考え方として，『機械学習品質マネジメントガイドライン』における開発プロセスの整理を紹介する。

AIシステムの開発では，前述した開発と運用を一体化したアプローチを通じてAIシステムの品質を確保していくことになる。一方で，AIの学習はデータに依存しているため，開発初期段階で成果物を予測することが難しく，また，開発者と利用者の間の認識にも違いが生じやすい。そのため，AIシステムの開発においては，新たな概念やアイデアの実現可能性を示すための検証プロセスであるPoC（Proof of Concept）が行われることが多い。また，AIシステムは，開発され実環境に導入された後も，精度の維持・向上のために，継続的に新たに取得したデータを用いた学習が行われる。

このような一般的なAIシステムの開発プロセスについて，産業技術総合研究所の発行する『機械学習品質マネジメントガイドライン』では，**図6**のように「PoCでの分析段階」「繰り返し型の開発段階」「DevOps的な運用改善段階」の3段階に分けて整理している。新技術導入にあたっての事前検証や，開発完了時の検証，運用開始後の変更の確認は海事業界においても重要であることから，このような開発段階の整理は，海事業界への親和性が高いと考えられる。

5. おわりに

本稿では，現在開発中の『船舶へのAIシステム活用に向けたテクニカルガイド -AIシステムの開発過程における品質管理-』に記載する内容として，AIの概要や，開発と運用を一体化したアプローチについて説明した。また，アプローチを海事業界の開発プロセスに適用する際の考え方の参考として，『機械学習品質マネジメントガイドライン』における開発プロセスの整理を紹介した。これらの内容はAIシステムの用途を問わない汎用的なものとなっているため，今後は，海事業界に適した開発プロセスを整理していく必要がある。

AIシステムについて海事業界では，「海事業界の人間がAIシステムを開発する」ケースと，「AIシステムの開発を担っている企業が海事業界向けのAIシステムを作る」ケースが出てくると考えられる。前者については，AIシステムの開発への理解が必要となり，後者については，海事業界の開発プロセスに対する理解が必要となる。現在開発中のガイドでは，両者のケースで気づきとなる情報を記載することを心掛けている。これらの内容が，海自業界におけるAIシステムの開発の一助となれば幸甚である。

参考文献

1)　JRCS：infoceanus, https://infoceanus.com/, 2024.3
2)　GROKE：Groke Pro,

https://www.groke-tech.com/en/products, 2024.3.
3) Orca AI：SeaPod, https://www.orca-ai.io/seapod/, 2024.3
4) 日本財団：無人運航船プロジェクト「MEGURI2040」未来の海を支える「無人運航船」の実用化を目指す, https://www.nippon-foundation.or.jp/what/projects/meguri2040, 2024.3
5) 日本財団：無人運航船プロジェクト「MEGURI2040」世界初の大型フェリーの無人運航実証，北九州市で成功, https://www.nippon-foundation.or.jp/who/news/pr/2022/20220117-66607.html,（2024年3月閲覧）
6) Hashimoto et al.：Development of AI-based Automatic Collision Avoidance System and Evaluation by Actual Ship Experiment, ClassNK technical journal, 2021
7) 総務省：令和元年版 情報通信白書｜AIに関する基本的な仕組み, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd113210.html, 2019
8) 佐藤文孝：Deep Learning概説―AIの核となる機械学習技術の最先端―, FUJITSU JOURNAL, 2018
9) 一色政彦：概念ドリフト(Concept drift)/データドリフト(Data drift)とは？, https://atmarkit.itmedia.co.jp/ait/articles/2202/21/news033.html,（2024年3月閲覧）
10) AI白書編集委員会：AI白書2023 生成AIのインパクトとAIガバナンス，角川アスキー総合研究所, 2023
11) PwC Japan：生成AIを巡る米欧中の規制動向 最前線 欧州「AI規則案」の解説, https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/generative-ai-regulation03.html,（2024年3月閲覧）
12) Tambiama Madiega：Artificial intelligence act, European Parliament Research Service, 2023
13) 独立行政法人情報処理推進機構：セキュリティ関係者のためのAIハンドブック, 2022
14) Ian J. Goodfellow et al.：Explaining and Harnessing Adversarial Examples, International Conference on Learning Representations ICLR, 2015
15) 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4版, 2023
16) ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology