

一般財団法人日本化学技術連盟

第36年度（2020年度）ソフトウェア品質管理研究会 成果発表  
研究コース5（人工知能とソフトウェア品質）ガイドラインチーム



## AIシステム品質評価ガイドライン具体化 におけるAI経験有無の影響明確化

研究員	池村 拓哉	（NTTコムウェア株式会社）
	佐藤 祥輝	（株式会社東光高岳）
	松本 充弘	（NTTテクノクロス株式会社）
主査	石川 冬樹	（国立情報学研究所）
副主査	栗田 太郎	（ソニー株式会社）
	徳本 晋	（株式会社富士通研究所）

# アジェンダ

---

- ・ チーム紹介
- ・ 研究の背景
- ・ 課題
- ・ アプローチ
- ・ 結果と考察
- ・ まとめ
- ・ 今後の課題
- ・ 一年間を振り返って

# ガイドラインチーム紹介

## 会社／名前／担当業務

## SQIP参加動機

NTTコムウェア株式会社

池村拓哉

- ・ 技術企画部門所属
- ・ プロジェクトマネジメント担当

- ・ AI技術を使用したシステム開発が増加しており、従来の品質評価手法では妥当性を判断する事がむずかしくなっているため情報収集を図りたい



株式会社東光高岳

佐藤祥輝

- ・ 研究開発部門所属
- ・ 技術調査および開発担当

- ・ AIを活用した製品を提供していくにあたって、その開発や検証・運用でどのように品質保証すべきか、情報を収集したい



NTTテクノクロス(株)

松本充弘

- ・ 品質保証部門所属
- ・ 担当事業部 品質保証担当

- ・ 徐々に増えつつあるA Iシステム開発に対し、品質管理プロセスを定義・提示する為の知識とノウハウ蓄積
- ・ A Iシステム開発プロジェクトへの、業界動向と情報の展開



**メンバ全員、AIシステムの開発経験ほとんど無し**

# 研究の背景（1 / 2）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

AIシステムの品質評価：**従来ノウハウが適用できない！**

近年、AI技術を用いたソフトウェアの実社会への導入が進んでいる

・自動運転、画像認識、不良品検査など

一方、品質測定・評価面では...

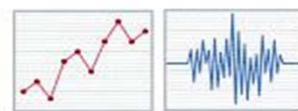
従来型システム  
(演繹的)

ロジックは  
人間が決定

実装

仕様書通りに実装されているか  
検証が可能  
品質管理ノウハウも多い

AIシステム  
(帰納的)



事実・  
データ

学習

事実・データをもとに学習して  
自動でパラメータ調整

学習結果が**ブラックボックス**  
であり、調整されたパラメー  
タが正しいか**検証が難しい**

AIシステムを  
開発したけど  
**品質**に問題は  
ないか？

AIシステムの品質保証技術が体系化・確立されているとは言い難い

## 研究の背景（2 / 2）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

最近、**AI品質評価ガイドライン**策定の動きあり

QA4AIコンソーシアム

「AIプロダクト品質保証ガイドライン（第2版：2020年2月1日）」

国立研究開発法人産業技術総合研究所（産総研）

「機械学習品質マネジメントガイドライン（第1版：2020年6月30日）」



### 当初のアクション（ガイドライン理解に向けて）

これらのガイドラインをAIシステム開発に適用すると  
システム品質が向上するか（適用前/後での品質差異）

- ①ガイドライン全項目の読み合わせ、**具体化**
- ②参考としたAIシステムの設計・評価に対して適用
- ③設計に適用した/しないで、品質が向上するか確認  
（実行/結果評価）



# 課題（1 / 3）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

①ガイドラインの具体化をやってみたが・・・



解釈が難しい！  
人によって解釈が違う！



## 課題（2/3）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

解釈が難しい・人によって解釈が違う理由は？

⇒チーム内で話し合ったところ、下記の事項が原因としてあがった

ガイドラインの記述が抽象的

AI品質ガイドラインの用語の意味がわからない

高レベルの基準をどうすればよいかわからない

公平性の比較ってどうすればいいの？

そもそも実現できる項目なのか？

工学的な検討・分析手法ってどんなことなの？



人によって解釈が違くと、品質評価の基準が変わってしまい  
本来満たすべき品質を満たせない可能性がある

解釈が難しい・人によって解釈が違う、**根本理由**は？

【仮定】

個人の知識・経験の違いで解釈の傾向に  
差異が発生するのでは？



【研究課題】

AIに関する知識・経験の有無によるガイドライン  
解釈の傾向を検証する

どうやって検証するか？





# アプローチ（1/2）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

## 【研究課題】

AIに関する知識・経験の有無によるガイドライン  
解釈の傾向を検証する



## 【アプローチ】

**アンケートによる調査**を実施

- ・ガイドライン各項目を解釈してもらう
- ・AI経験有無で回答、**着眼点**の差異の傾向があるか確認



# アプローチ（2/2）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

## アンケート内容

対象ガイドライン	産総研ガイドライン データ被覆性(lv1)に限定
対象AIシステム	クレジットカードデフォルト 予測モデル(※1)
回答形式	自由記述形式
対象者	AI経験者(8名)・未経験者(14名) →計22名から回答
評価方法	キーワードの出現頻度(※2)

- ・理解しやすい
- ・具体化しやすい

- ・理解しやすい
- ・公開されている

- ・回答者の考えを知りたい  
(選択式で回答誘導したくない)

- ・研究会参加者と  
その会社内部の協力者

(※1) 対象AIシステム

kaggle, 「Default of Credit Card Clients Dataset」

<https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset>

(※2) 評価方法

ガイドライン解釈内容の回答において、**AI経験者に共通する/特徴的な**キーワードを抽出し、その出現回数をAI経験者/未経験者別に集計・比較



# 結果と考察（1 / 4）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

## アンケート項目（ガイドライン項目内容）

項番	項目内容（データ被覆性 lv1）
①	テスト用データセットの取得源や方法を検討し，応用の状況に対して偏りが無いことを期待できるようにすること
②	各ケース毎に，元データから偏りのないサンプル抽出などを行い，偏りが無いことを期待できるようにすること
③	これらの偏りを入れないために行った活動について，記録を行うこと
④	分析した各ケースについて訓練用データおよびテスト用データが十分に存在することを，訓練フェーズやバリデーションフェーズなどで確認すること
⑤	ケースに対して訓練用データが十分に取得できない場合には，網羅基準を見直して緩めた上で，当初の基準に照らして個別にシステム結合テストなどで確認すべきことを記録しておくこと

## 結果と考察（2 / 4）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

### アンケート結果（キーワードの傾向）

項目（被覆性 lv1） （ポイント抜粋）	AI未経験者に多い	AI経験者に多い	差異がない
①テストデータ セット取得源 ・方法	特になし	母集団分布と近似	合成データ利用 データの信頼性 複数のデータ入手元 大量データ利用
②元データからの 偏りのない抽出	各ケースのパターン別 件数定義	偏り是正（データ合成） 偏り是正（抽出件数・方 法）	各ケースの確認
③上記活動の記録 内容	データ・件数定義の記 録 データ取得方法の記録	データ加工手法の記録 加工前後の記録	特になし
④データ十分性の 確認	特になし	スコープ（要求仕様）との 比較 データの十分性確認 （検証・評価）手法	入手元 学習 （統計的）指標
⑤データ不足時の 活動記録	特になし	基準見直し経緯 見直し前後の結果 ステークホルダとの合意 見直し影響（手戻りなど）	データ件数が少ない レアケース データ偏り

## 結果と考察（3 / 4）

背景

課題

アプローチ

結果と考察

まとめ

今後の課題

### 各項目別の考察

項目（被覆性 lv1） （ポイント抜粋）	差異に対する考察 （AI経験者特有の着眼点）
①テストデータ セット取得源 ・方法	<ul style="list-style-type: none"><li>・ 現実のデータと近似の状態である必要性</li><li>・ 学習データによる振る舞いについての知識</li></ul>
②元データからの 偏りのない抽出	<ul style="list-style-type: none"><li>・ 学習させるためのデータは網羅的に収集できないことが多いこと</li><li>・ 不十分であれば学習に必要なパターンの観点で合成などの手法により作成すること</li></ul>
③上記活動の記録 内容	<ul style="list-style-type: none"><li>・ データ加工がAI学習内容に影響を与えること</li></ul>
④データ十分性の 確認	<ul style="list-style-type: none"><li>・ AIシステム品質が学習データに強く依存していること</li></ul>
⑤データ不足時の 活動記録	<ul style="list-style-type: none"><li>・ 従来型システムとは異なり、AIシステムの品質基準はお客様と同意して勧めていく必要があり、その基準緩和を行う際にはお客様と再度品質基準を合意すること</li></ul>

## 全体の考察

### AI未経験者の傾向

AIに関して一般的な内容を回答している

例：データを大量に準備する、収集時の状況を記録する  
→AI経験者にとっては当たり前すぎて回答していないのでは？

### AI経験者の傾向

各種評価指標や機械学習固有の手法を考慮している

例：母集団との近似・偏り是正、不足データの合成

### 共通した傾向

従来型システム開発時にも意識すべき観点は確認している

例：各ケースの確認、レアケースなど



## 【課題】

AI品質ガイドライン：抽象的、利用には具体化が必要で  
品質基準のばらつきが生まれる  
⇒**解釈の差異を把握**する

## 【アプローチ】

AI経験の有無による差異の確認が必要  
⇒アンケート実施・評価

## 【わかったこと】

AI経験の有無で差異の傾向あり  
⇒AI経験者は各種統計量やデータ合成といった  
**指標や機械学習固有の手法を考慮**している

## 【結果の活用】

統計量やデータ合成など、具体的な観点をリスト化  
AI経験者をレビューに加える



以下事項の確認は今後の課題と考える

- 分析対象の拡大
  - ・ガイドライン項目の対象範囲拡大
  - ・AI経験有無以外の要因
- アンケート母数・方法の見直し
  - ・アンケート母数の拡大
  - ・アンケート回答方法
  - ・アンケート集計方法
- 結果の活用
  - ・AI未経験者への教育
  - ・ノウハウの共有(資料化等)



ガイドラインを十分な精度をもって具体化できるように！



# 1年間を振り返って

---

- ・私自身もAIシステムに関する知見が殆どない中、研究テーマの検討やその後の取り組みを手探りで進めて行く中で、指導頂いた皆様やチーム関係者の方のご支援、ご協力を頂く事で何とか発表の場に漕ぎつけました事感謝いたします。
- ・この研究会を通じて得られた知見や経験を自社に持ち帰り、引き続き本テーマ含めたAI品質に対する取り組みを進めていき業務に活かせる様にしたいです。



- ・直接顔を合わせる機会も無いなかでの共同研究は戸惑いも多かったですが、Web会議を頻繁に行って密な連携を取れたのでいい経験になりました。
- ・AIの品質に関して様々なことが学べましたので、今後は社内にも展開していきたいです。



- ・当初はAIシステムに関する知識が足りずに苦労しましたが、ガイドラインを読み込み、GoogleColaboratory や Kaggle などを活用する事で理解が進むと、久しぶりのプログラミングなども有り、楽しくなってきました。今後も勉強を進め、知見を増やしたいと思います。
- ・時間に追われながらでしたが、社外の方々との共同作業は貴重な体験でした。



## 謝辞

---

石川主査、栗田副主査、徳本副主査、1年間  
ご指導ありがとうございました。

日科技連事務局の皆さま、  
ご対応ありがとうございました。

本実験にご協力いただいた研究コース5の  
メンバーおよび関係者の皆さま、  
ありがとうございました。

研究コース5 GLチーム 一同

---

ご清聴ありがとうございました