

第36年度 研究コース5「人工知能とソフトウェア品質」(チームGL)

AI 品質マネジメント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 (エヌ・ティ・ティ・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)

研究概要

AI システム開発におけるソフトウェア品質評価に関しては、従来型の品質保証手段が利用できず、その品質保証技術が体系化・確立されているとは言い難い状況にあったが、昨年より幾つかの組織において品質管理ガイドラインが発表され始めている。しかしガイドラインが抽象的であるため、個人によるガイドライン解釈の幅が大きくなる傾向にある。そこで本研究では、ガイドライン利用者の AI 開発経験有無に着目し、それらの差異により解釈の方向性が異なる傾向にあると仮定した。産総研のガイドラインを基にアンケートを実施し、22 名から回答を得た。その結果、AI システムの学習に利用するデータに対し、AI 未経験者では具体的な加工手法や統計的手法を交えた観点が漏れてしまう傾向にあり、AI システムの品質向上を図るためには AI システム経験者を参画等により前述の観点の漏れを防ぐ事が重要であると分かった。

1. 研究の背景・現状

近年、様々な分野において AI 技術の利用が進展しており、AI 技術を用いたソフトウェア開発が行われ実社会への導入が進んでいる。その一方でソフトウェア品質評価の面に関しては、計算や判断をおこなうためのモデルやアルゴリズムを人が決定する従来の演繹的システムに比べると、モデルやアルゴリズムを学習データから獲得生成する帰納的システムである AI システムでは、従来型の品質保証手段が利用できず、その品質保証技術が体系化・確立されているとは言い難い状況にある。[1]

2019 年頃から AI システム開発における品質管理ガイドラインについて、幾つかの組織において策定の取り組みがなされており、例として QA4AI コンソーシアムから「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) から「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)」が発表されている。[2][3]

しかしこれらのガイドラインは出来るだけ広い範囲のシステムに適用できるように、汎用的な記述内容になっており、実際の利用に即しては案件毎に内容を選択・具体化して利用することが想定されている。そのため、ガイドラインの内容を実際の AI システムで利用するためには、対象となる AI システムの要求事項等を考慮したうえで内容の具体化を行ない、対象の AI システムに適用し品質評価をおこなう事となるが、当チームメンバ (AI 技術の経験なし) がガイドラインの適用を試みたところ、各々の解釈が分かれる結果となった。これは、具体化の指針が明確に定められていないため、個人のもつ知識や経験により具体化の内容に大きく差異が発生したためと想定される。そこで本研究では、産総研のガイドラインを対象に、その具体化にあたって AI システム開発経験やその他知識・経験の有無が、ガイドライン項目の具体化作業時にどのような影響を及ぼすかに関して調査をおこなう。

2. 研究課題

産総研ガイドラインの項目は汎用的・抽象的であるため、実際の AI システムへ適用するには、各ガイドライン項目の具体化が必要である。しかし、具体化のための指針が明確にはないことから、作業者の AI システムに関する知識・経験の有無等によりガイドラインの解釈に差が生じる。具体化した結果が人によりばらつきが出る事により、本来確認すべき事項が漏れる事で、担保すべき品質が満たせなくなる事が問題と考えた。

その解釈の差を把握する事が AI システム品質の向上に対して有効と捉え、これによりガイドラインの利用をより促進出来ると考えチームの研究課題とした。

3. 課題解決のためのアプローチ

3. 1 アンケートによる調査

ガイドラインの解釈差異については、AI システムの知識や経験の有無による差異があると考えた。また調査対象を広くとる事と合わせ、ガイドラインの各項目の解釈に関するアンケートを以下の方針で実施した。

①産総研ガイドラインを採用

AI 品質管理に関するガイドライン資料は何点か発表されており、代表的なものとして前述の QA4AI コンソーシアムのものと産総研のものがあるが、QA4AI は開発プロセス全体を5つの視点でみているところ、産総研は開発プロセスの流れに従って特性をみていることから、特定の開発プロセスと特性に絞ったアンケートが実施しやすいため、産総研ガイドラインを対象として採用した。

②ガイドライン観点を絞って実施

産総研ガイドラインは、全体構造として「内部品質」「外部品質」「利用時品質」の3つの機械学習要素に分かれているが、アンケートは従来型ソフトウェア品質の考え方に比較的近い「内部品質」とし、またその中の6つの観点中「データセットの被覆性」の1v1を対象として選択した。これはアンケート回答者に AI 未経験者も含まれるため、ガイドラインの品質要求事項が比較的具体的な項目や1v1を選んだ。

③AI システムの指定

回答者が個々に異なる AI システムを想定し、それを前提としてアンケートに対して回答を行う事で解釈のばらつきが発生する事を防ぐため、対象の AI システム概要をアンケート時に併せて提示した。

3. 2 アンケート内容

アンケート実施者の恣意性が回答に影響する事を避けるため、ガイドラインの要求事項に対する質問を作成し（自由記述形式での回答）、回答から要求事項の理解・解釈の差を比較・評価する事とした（アンケート内容は付録を参照）

また回答の集計方法は、AI 経験者の回答に記載されている AI 開発時に考慮すべきと考えた観点をキーワードとして抽出し、アンケートにどの程度の頻度で出現しているかを、AI 経験者、未経験者別に集計した。

①対象となるガイドライン項目

産総研ガイドラインのうち内部品質の「データの被覆性」1v1（要求事項5項目）に限定した。

表 1 内部品質：データの被覆性 1v1 5項目

項目	要求事項の説明内容
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。

②対象 AI システム

アンケートを実施する際の前提として回答者へ提示する AI システムについては「クレジットカードデフォルト予測モデル」を指定した。[4] これは kaggle 上で公開されているものであり、プログラムやデータセット内容が容易に入手可能である事、システムとして特殊性が少ない事から選択している。

③収集情報

各要求事項に対する自由記入形式のアンケート回答用紙を作成した。またその他情報として、回答者の属性情報（AI 業務経験有無・年数、業務分野、役割、経験開発手法、等）をあわせて収集し、これらの属性による解釈の傾向が見られるか調査した。

④アンケート対象者

AI システム経験のある有識者を母集団に含めるため、第 36 回 SQiP 研究会の同研究コース（AI とソフトウェア品質）の参画者をまず被験者として依頼し、またシステム開発・品質保証の経験者も母集団に広く含めるために、本チーム研究員の社内組織関係者に対しても依頼した。

⑤アンケート回答の評価方法

回答が選択枝ではなく自由記入のため、各アンケートの回答から AI 経験者/未経験者でのキーワード出現頻度により考慮有無の比較をおこなう。

4. 結果

アンケートへの回答状況として、全回答者 22 人のうち AI 経験者は 8 人、AI 未経験者は 14 人であった。アンケート各要素項目のキーワードに対する集計結果を図 1～図 5 に示す。また、アンケート回答者の属性情報の分布は付録に記載する。

<項目 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>

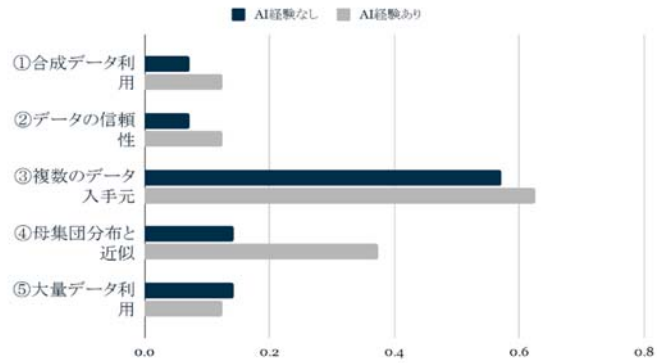


図1 項目1回答のキーワード含有率

項目1より、以下のキーワードで回答者のAI経験有無により差異が出ている事がわかった。

- ・④母集団の状態に近い(通常/レアケースの内容、頻度)

＜項目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＞

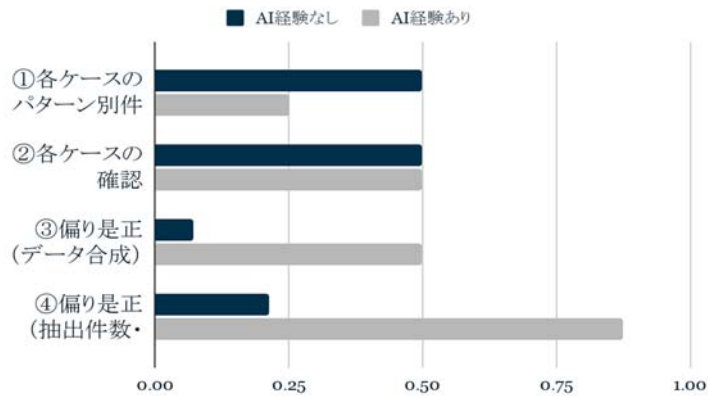


図2 項目2回答のキーワード含有率

項目2より、以下のキーワードで回答者のAI経験有無により差異が出ている事がわかった。

- ・①各ケースのパターン別に必要な件数を定義：AI未経験が優位
- ・③データ偏りを、データ合成では正、収集・抽出方法では正：AI経験者が優位

＜項目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＞

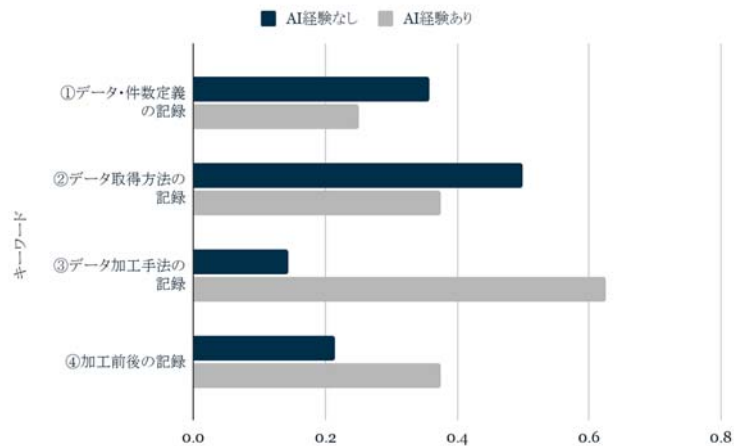


図3 項目3回答のキーワード含有率

項目3より、以下のキーワードで回答者のAI経験有無により差異が出ている事がわかった。

- ・①データ件数定義の記録、②データ取得方法の記録：AI未経験者が優位
- ・③データ加工手法の記録、④データ加工前後のデータセットの記録：AI経験者が優位

＜項目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＞

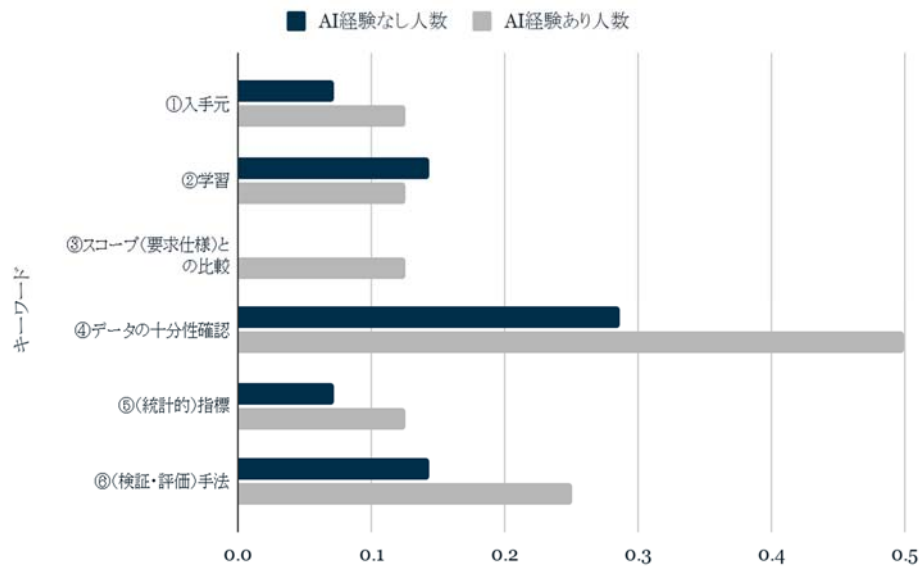


図4 項目4回答のキーワード含有率

項目4より、以下のキーワードで回答者のAI経験有無により差異が出ている事がわかった。

- ・③スコープ（要求仕様）との比較、④データの十分性、⑥検証・評価手法：AI経験者優位

＜項目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＞

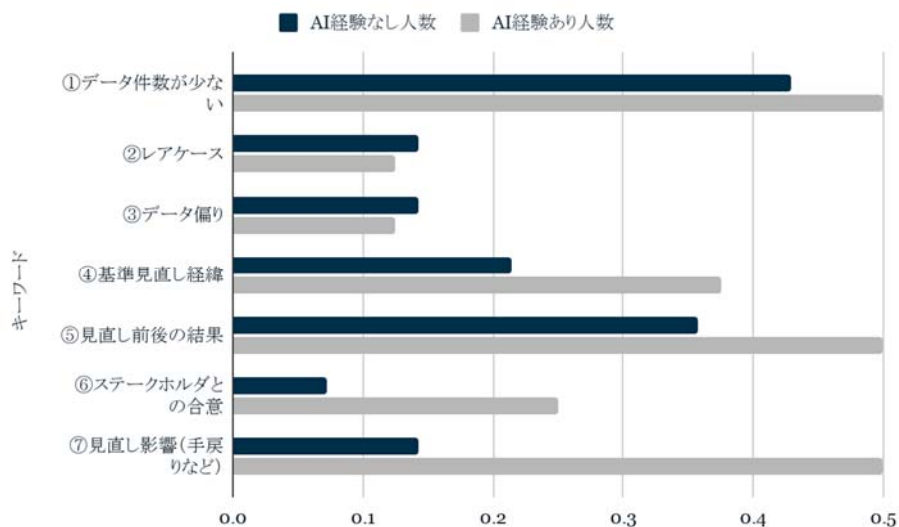


図5 項目5回答のキーワード含有率

項目5より、以下のキーワードで回答者のAI経験有無により差異が出ている事がわかった。

・④基準見直し経緯，⑤見直し前後の結果，⑥ステークホルダとの合意，⑦見直し影響（手戻りなど）：AI経験者で優位

5. 考察

5.1 アンケート結果の考察

表1の各項目について、アンケート結果からデータセットの被覆性に関して言える事として下記の様に考察をおこなった。

・項目5.の結果から、テスト用データセットの取得源に関しては、AI未経験者はデータを大量に収集する事に言及している一方、AI経験者はそれに加えて母集団の状態に近いか(通常/レアケースの内容、頻度)に言及している。これはAI経験者が現実のデータと近似の状態である必要性を重視している一方、AI未経験者は学習データによる振る舞いについての知識がない事から、大量に集めることを重視したためと考えた。

・項目2.の結果から、元データから偏りなく抽出を行う際には、AI未経験が各ケースのパターン別に件数を定義する事を指摘している一方、AI経験者側は定義した件数の収集が難しい場合にデータ合成や抽出方法での是正に言及している。未経験者は事前に決めた処理パターンを網羅する様なデータ収集自体を重視している一方、経験者は学習させるためのデータは網羅的に収集できない事が多く、不十分であれば学習に必要なパターンの観点で合成等の手法により作成する事も考慮に入れているためと考えた。

・項目3.の結果から、データの偏りを入れないために行った活動記録については、AI未経験者ではデータ件数やデータ取得方法といった収集に関する記録に言及している一方、AI経験者はデータ加工手法やデータ加工前後のデータセットの記録に言及している。これは、AI経験者はデータ加工の活動がAI学習内容に影響を与える事を理解しているが、AI未経験者はその考え方に至らないためと考えた。

・項目4.の結果から、訓練用・テスト用データの十分性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事に関して、AI経験者はAIシステムに要求される仕様(スコープ)に対してデータが十分揃っているかや、検証・評価手法についての言及が未経験者に対して多い。これは、AI経験者はAIシステム品質が学習データの内容に強く依存していることを理解しているが、AI未経験はその知識が無く考え方に至らないためと考える。

・項目5.の結果から、訓練用データが十分に取得できない場合の基準緩和とテスト時に確認すべき事項の記録では、AI未経験者ではデータ件数や基準見直しの経緯、見直し前後の値の様に、見直しに関する作業そのものの記録に言及しているが、AI経験者ではそれらに加えてステークホルダとの合意や手戻り等の見直し影響にも言及している。これは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進めていく必要があり、その基準緩和をおこなう際にはお客様と再度品質基準を合意する事となるため、AI経験者側はその経験からステークホルダとの合意・見直しのキーワードが、AI未経験者と比較して多く出てきているためと考える。

5.2 全体としての考察

それぞれのアンケート項目への回答に対する考察から、アンケート回答の全体的な傾向として以下の様に考えた。

・AI未経験者は、データを大量に準備する、収集時の状況を記録するといった、AIシステムに限らず従来型システム開発でも考慮を行うべき様な内容を主に挙げている一方、AI経験

者はデータが母集団に近似の状態か、統計的に十分か、十分性や偏りが満たせない場合に不足データを合成によって確保するという様に、具体的にどの様な対応策を取るかの観点に基づいて挙げている。

- ・大量データを収集する、収集時の記録を残すといった AI 未経験者側の回答で優位な項目については、AI 経験者にとっては当然の活動であるため挙げていない。
- ・AI 経験者/未経験者で差異があまりみられなかったものは、従来型システム開発時にも意識すべき様な観点であった。

5. 3 本調査結果の活用について

各社で実際の AI システム開発案件への利用を目的として開発部門・品質管理部門で AI ガイドラインの具体化をおこなう際に、考慮すべき観点の漏れを防ぐためには AI 経験者の参画が重要である事を周囲に理解して貰うための、会社や組織への説明資料として本調査結果を活用できると考える。

6. おわりに

6. 1 まとめ

本研究では、AI 品質評価ガイドラインを具体的な案件に対して適用しようとした場合、その作業者の AI システム経験有無によりガイドライン解釈の差異が大きく、AI 未経験者では AI 経験者と比べて必要な考慮事項が漏れてしまう傾向にある事が分かった。AI システム品質の向上を図るために、学習・テストデータ設計等でガイドラインを元にレビュー観点を具体化する際には、AI システム経験者を参画させる事で考慮すべき観点の漏れを防ぎ品質を向上する事が出来ると考える。

6. 2 今後の課題

今回、産総研ガイドラインの具体化に際しては、AI 未経験者は従来型開発の観点をを用いるが、AI 経験者はそれに加えてデータ準備に不足があった場合の具体的な対応まで考慮した観点をを用いている事が明確になった。ただし、以下に示すような事項の確認が不十分でありさらなる調査が必要と考える。最終的には、ガイドラインの全要素に対して十分な精度をもって具体化出来る事がゴールであるが、そこまでを実現するためには長期的な取り組みを継続する必要がある。

・ガイドライン項目の対象範囲拡大

今回、アンケートの対象範囲が限定的（内部品質の「データの被覆性」1v1（要求事項 5 項目））であったため、調査結果の信頼性を向上させるためにも、他のガイドライン項目・1v でも今回の考察と同様の傾向にあるのか、同様のアンケートによる検証の取り組みが必要であると考え。

・アンケート母数の拡大

今回のアンケートは回答母数が 22 件であったため、定量的な分析精度をあげるためには、対象を拡大してアンケートを実施する必要があると考える。

また今回アンケートの中で収集した業務分野、経験年数、年齢等の属性情報については、回答母数が少なかった事もあり明確な傾向としては確認出来なかったため今回の分析軸には利用しなかったが、アンケート母数を増やすとその辺りの傾向が見えて来る可能性もあると考える。

・アンケート回答方法

今回のアンケートは自由記述形式であったため、回答者にとって回答方法の基準が無く記載

レベルがまちまちであったため、集計時の解釈が難しかった。そのため、自由記述であっても回答粒度を揃えやすく集計がもっと容易になるような工夫をする、あるいは回答を選択式として選択項目内容を工夫する等の取り組みを行う事も考えられる。

- ・アンケート集計方法

今回の集計方法は、AI 経験者の回答中に含まれる観点をキーワードとして回答への出現頻度を比較しているため回答者の文章表現に依存される。アンケート以外の手法も用いたより適切な集計方法で本研究テーマに取り組む事も必要であると考ええる。

- ・AI 未経験者への教育

AI 経験者の参加による解決だけでは人材的な不足を補う事が難しいため、AI 未経験者の知見を AI 経験者へと近づけるための施策を検討する事も、今後 AI 案件が増加していくと想定される状況においては必要な取り組みであると考ええる。

- ・AI 経験有無以外の要因

今回の結論は AI 経験者の参加による解決を提示しているが、AI 経験者が参画しても解決できない問題の有無やその内容については明らかに出来ていないため、この問題に取り組む事でより問題の本質を捉える事が出来ると考える。

参考文献

[1] 石川 冬樹 (NII) Web サイト 「AI 時代における品質保証のチャレンジ～機械学習の難しさと (AI による) テスティング」

<http://research.nii.ac.jp/~f-ishikawa/work/1807-ESTIC18-AI+Testing.pdf>

[2] CPSEC, 人工知能研究センター, 「機械学習品質マネジメントガイドライン 第 1 版」

<https://www.cpsec.aist.go.jp/achievements/aiqm/AIQM-Guideline-1.0.1.pdf>

[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 2020.02 版」

<http://www.qa4ai.jp/QA4AI.Guideline.202002.pdf>

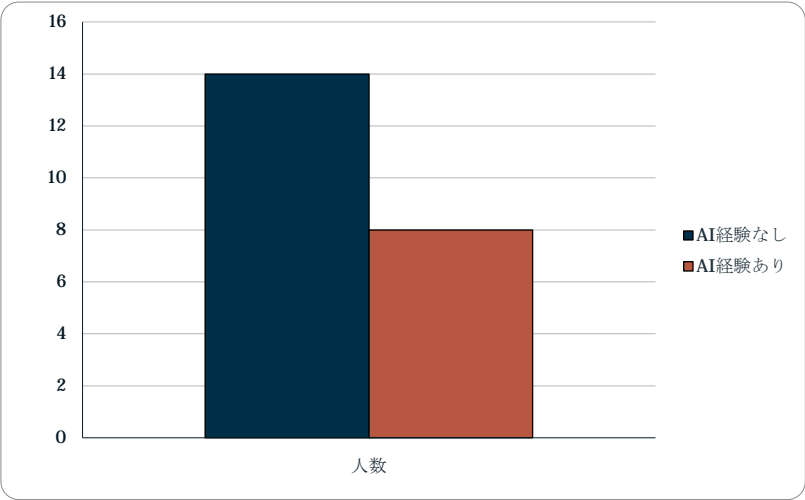
[4] kaggle, 「Default of Credit Card Clients Dataset」

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

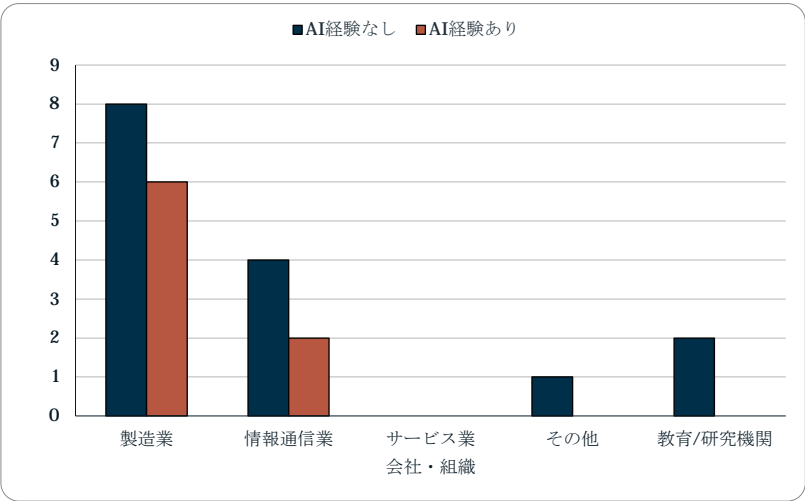
【所属情報の傾向について】

・今回のアンケートで収集した所属情報は、AI経験有無を除く他の情報（年齢、所属、役割等）については収集母数が少ないことも影響し、明確な傾向はみられなかったため、論文本文でのアンケート結果考察には使用していない。

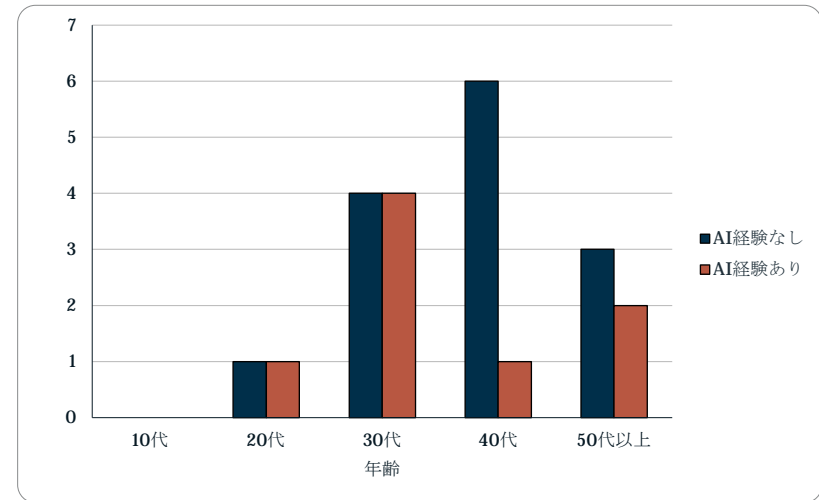
アンケート回答数	人数	AI経験なし	AI経験あり
		14	8



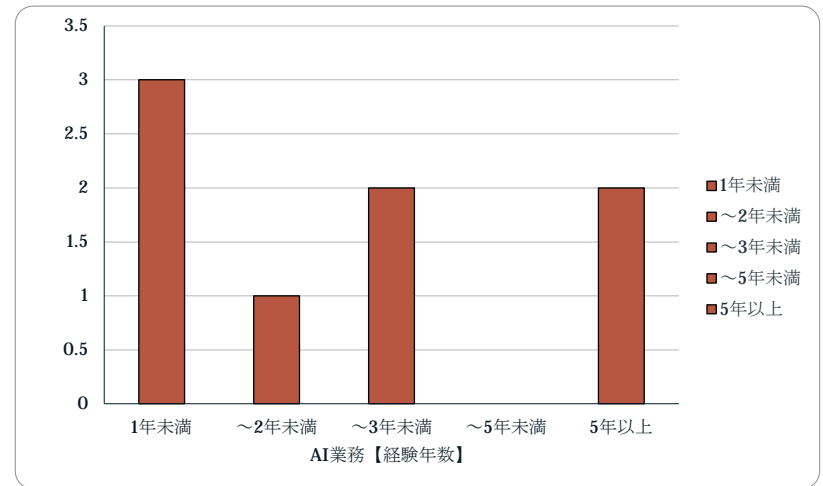
会社・組織		AI経験なし	AI経験あり
民間	製造業	8	6
	情報通信業	4	2
	サービス業	0	0
	その他	1	0
	教育/研究機関	2	0
	その他	0	0



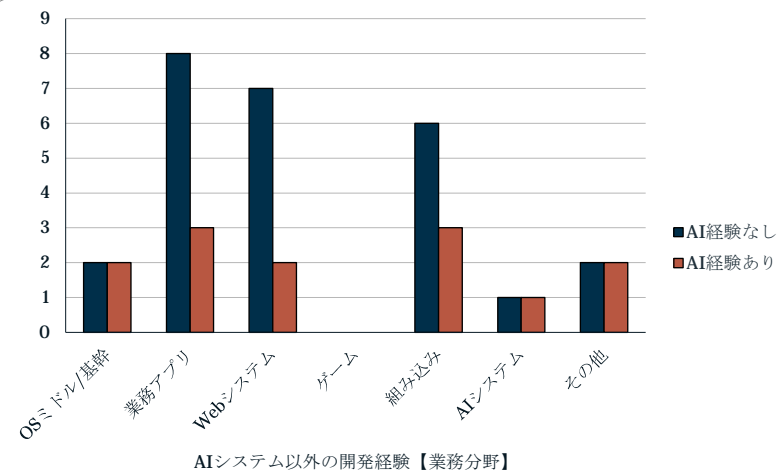
年齢		AI経験なし	AI経験あり
	10代	0	0
	20代	1	1
	30代	4	4
	40代	6	1
	50代以上	3	2
	その他/無回答	0	0



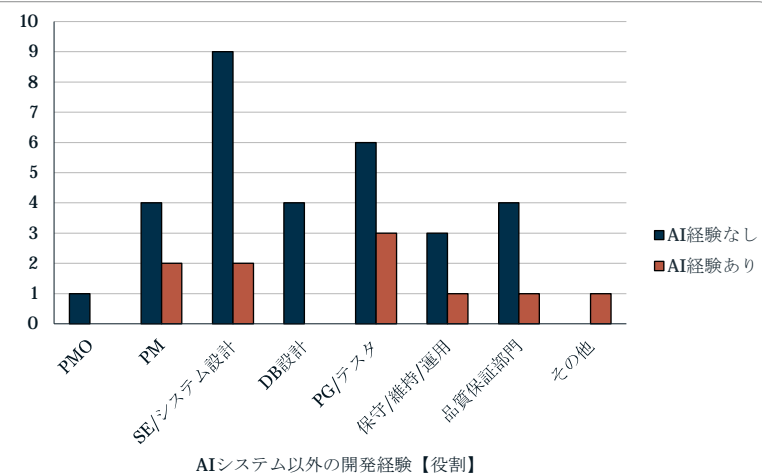
AI業務【経験年数】		AI経験あり
	1年未満	3
	～2年未満	1
	～3年未満	2
	～5年未満	0
	5年以上	2



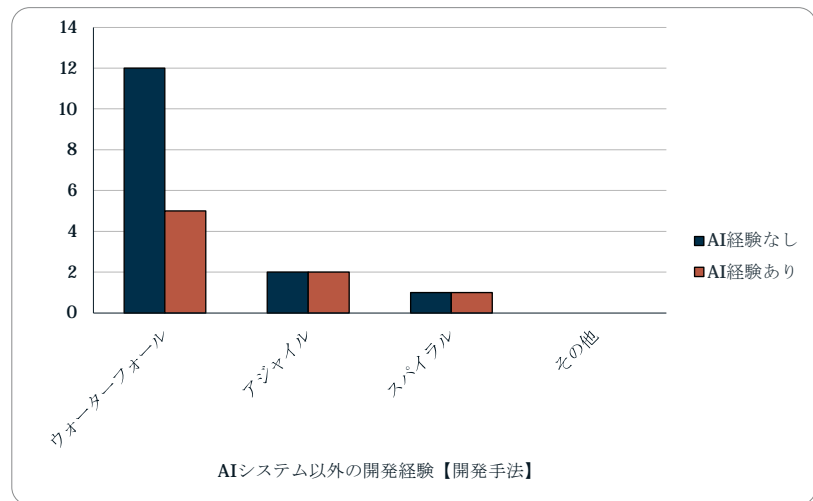
AIシステム以外の開発経験		AI経験なし	AI経験あり
業務分野	OSミドル/基幹	2	2
	業務アプリ	8	3
	Webシステム	7	2
	ゲーム	0	0
	組み込み	6	3
	AIシステム	1	1
	その他	2	2



AIシステム以外の開発経験		AI経験なし	AI経験あり
役割	PMO	1	0
	PM	4	2
	SE/システム設計	9	2
	DB設計	4	0
	PG/テスト	6	3
	保守/維持/運用	3	1
	品質保証部門	4	1
	その他	0	1

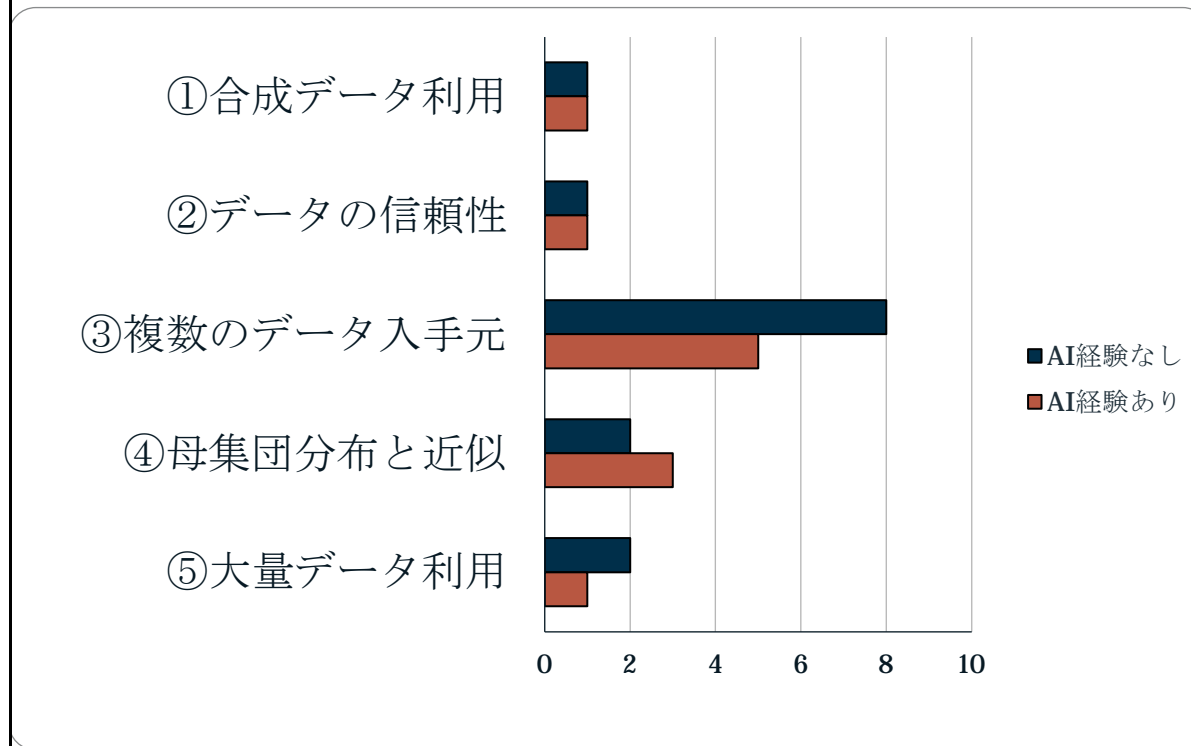


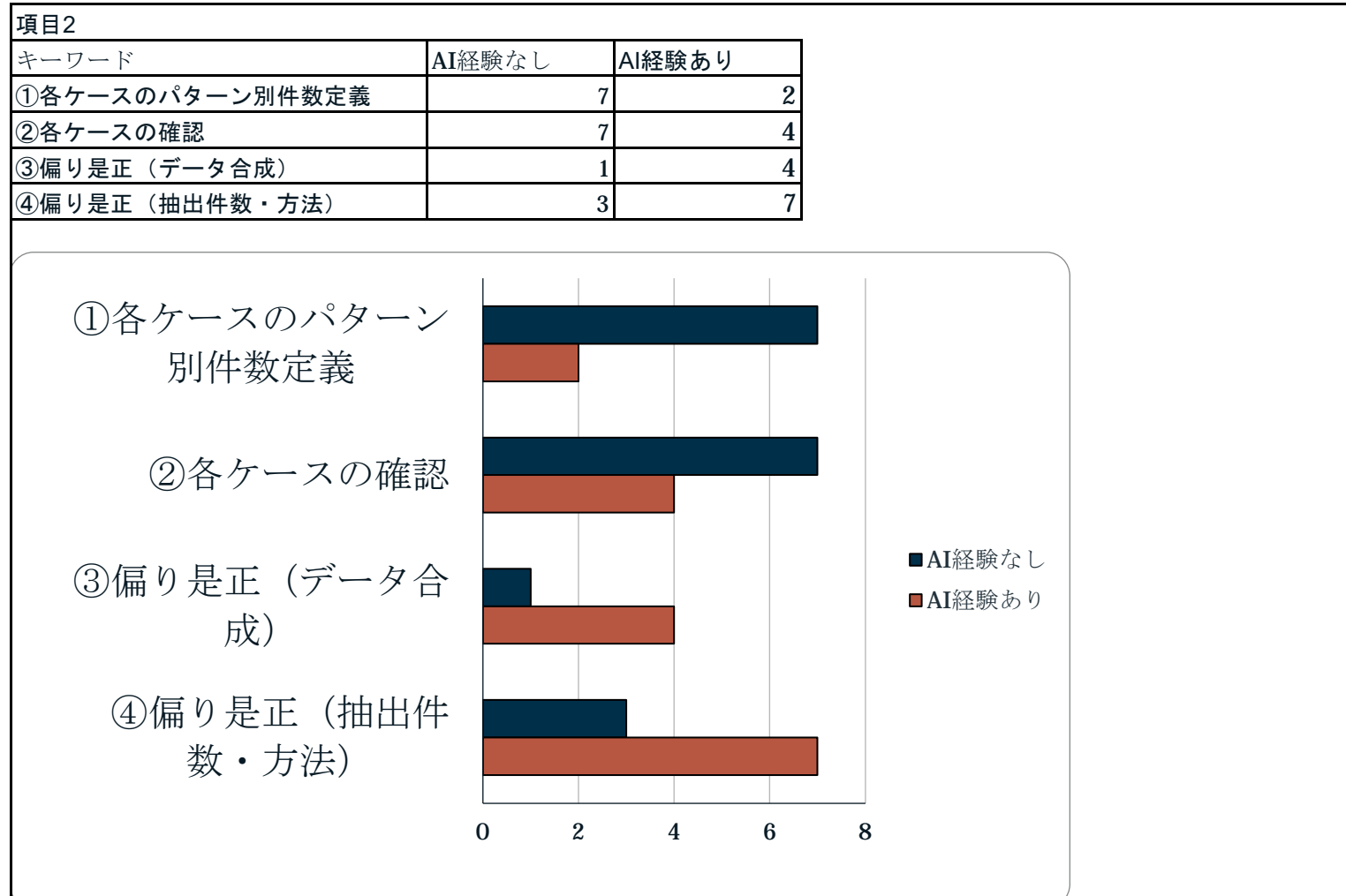
AIシステム以外の開発経験		AI経験なし	AI経験あり
開発手法	ウォーターフォール	12	5
	アジャイル	2	2
	スパイラル	1	1
	その他	0	0



項目1

キーワード	AI経験なし	AI経験あり
①合成データ利用	1	1
②データの信頼性	1	1
③複数のデータ入手元	8	5
④母集団分布と近似	2	3
⑤大量データ利用	2	1





項目3

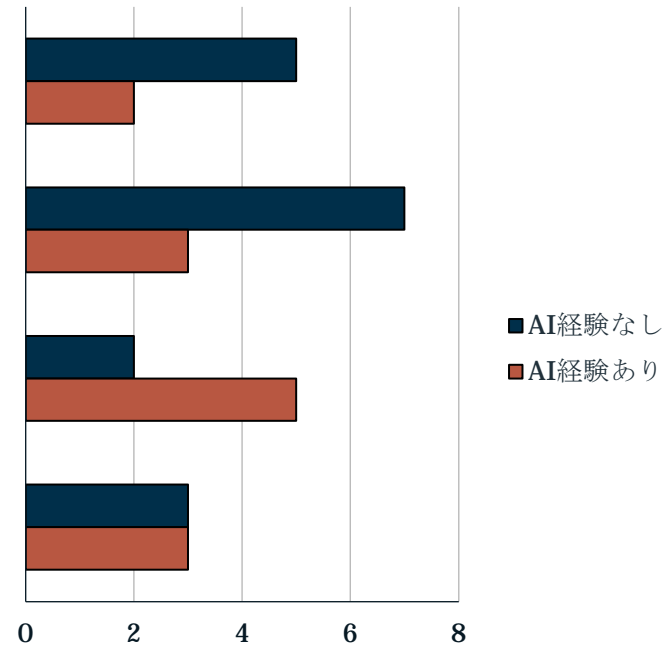
キーワード	AI経験なし	AI経験あり
①データ・件数定義の記録	5	2
②データ取得方法の記録	7	3
③データ加工手法の記録	2	5
④加工前後の記録	3	3

①データ・件数定義の
記録

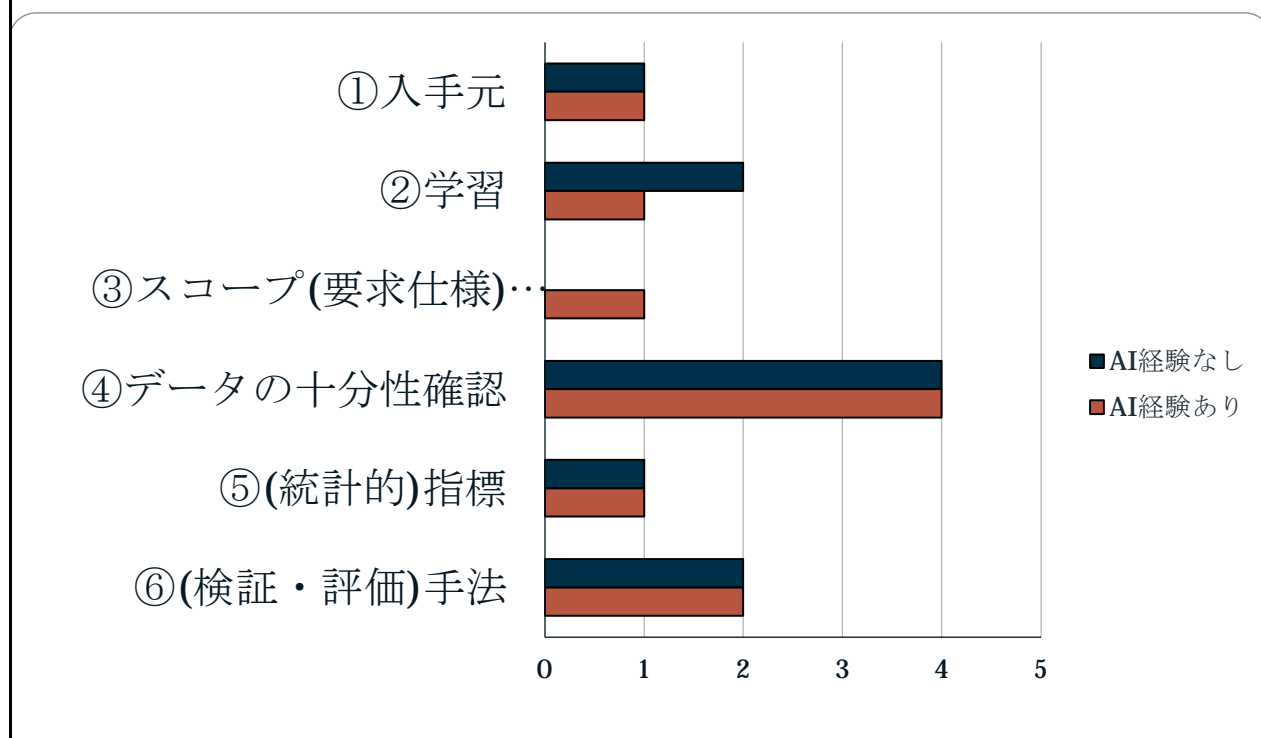
②データ取得方法の記
録

③データ加工手法の記
録

④加工前後の記録



項目4		
キーワード	AI経験なし	AI経験あり
①入手元	1	1
②学習	2	1
③スコープ(要求仕様)との比較	0	1
④データの十分性確認	4	4
⑤(統計的)指標	1	1
⑥(検証・評価)手法	2	2



項目5

キーワード	AI経験なし	AI経験あり
①データ件数が少ない	6	4
②レアケース	2	1
③データ偏り	2	1
④基準見直し経緯	3	3
⑤見直し前後の結果	5	4
⑥ステークホルダとの合意	1	2
⑦見直し影響(手戻りなど)	2	4

①データ件数が少ない

②レアケース

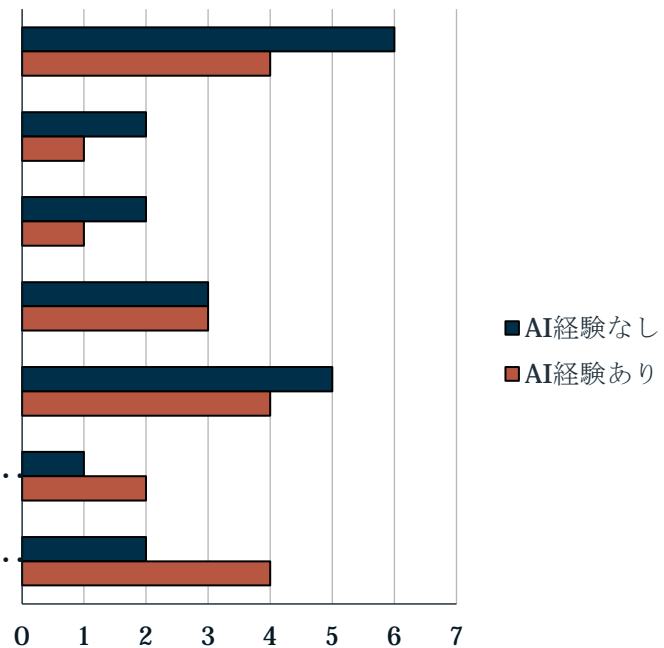
③データ偏り

④基準見直し経緯

⑤見直し前後の結果

⑥ステークホルダとの合意

⑦見直し影響(手戻りなど)



回答者の属性情報記入のお願い

アンケート回答集計時の重要な情報となりますので、差しさわりのない範囲で記載をお願いします。

ただし、「AIシステムに関する業務経験」有無については、重要な集計情報となりますのでご記載をお願いします。

※下記黄色塗りの記入欄へ、①やaの選択肢を記載願います(9行目のみ自由記述となります)

現在所属されている会社・組織	【	】	①民間企業（a.製造業、b.情報通信業、c.サービス業、z.その他）、②教育/研究機関、⑨その他
年齢	【	】	①10代、②20代、③30代、④40代、⑤50代以上、⑥その他(無回答含む)
AIシステムに関する業務経験	【	】	①あり、②なし
→「あり」の方			
・経験年数	【	】	a.1年未満、b.～2年未満、c.～3年未満、d.～5年未満、e.5年以上
・業務領域	【	】	※自由記述
AIシステム以外の開発経験	【	】	①あり、②なし
→「あり」の方			
・業務経験（分野）	【	】	a.OSミドル/基幹業務システム、b.業務アプリ、c.Webシステム、d.ゲーム、e.組み込み、f.AIシステム、z.その他
・業務経験（役割）	【	】	a.PMO、b.PM、c.SE/システム設計、d.DB設計、e.PG/テスト、f.保守/維持/運用、g.品質保証部門、z.その他
・主な開発手法	【	】	a.ウォーターフォール、b.アジャイル、c.スパイラル、z.その他

アンケート記入シート

品質管理ガイドラインには様々な要求事項があります。その中でも内部品質の「被覆性」に関する5つの要求事項に関する各質問にご回答願います。

下記の「要求事項」を読み、「質問事項」に対する回答を、「回答欄」へ記述願います。

質問に対し、何を回答すべきか（問われているか）分からない等あれば、その旨を回答欄に記載して下さい（～の様に考えたが…など、お考えをできるだけ詳しく記載頂けると助かります）

項目	要求事項	質問番号	質問項目	回答欄
1-1	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。	1	偏りをなくするためには、データをどこから、どのように揃えれば良いと考えますか？	
		2	上記を実施する際には、どのような不明点/課題/問題点があると考えますか？	
		3	何かコメントや注意点あれば、記載して下さい	
1-2	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。	4	「偏りのないサンプル抽出」を可能とするためには、どのような取り組みをすればよいと考えますか？	
		5	上記を実施する際には、どのような不明点/課題/問題点があると考えますか？	
		6	何かコメントや注意点あれば、記載して下さい	
1-3	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。	7	どのような項目をどのように記録する必要があると考えますか？	
		8	上記を実施する際には、どのような不明点/課題/問題点があると考えますか？	
		9	何かコメントや注意点あれば、記載して下さい	
1-4	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。	10	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確認するためには、何を実施しますか？	
		11	上記を実施する際には、どのような不明点/課題/問題点があると考えますか？	
		12	何かコメントや注意点あれば、記載して下さい	
1-5	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。	13	基準を見直す（妥協する）必要があるのは、どのような時だと考えますか？	
		14	基準を見直した事に対し、試験において確認・記録すべき事は何かと考えますか？	
		15	上記を実施する際には、どのような不明点/課題/問題点があると考えますか？	
		16	何かコメントや注意点あれば、記載して下さい	

本シートは、今回アンケートの対象となる、機械学習モデル（AIモデル）とそれに使用するデータ要素、

＜対象のAIモデルについて＞

XX銀行では、クレジットカードユーザーの債務不履行による損益が毎年発生しています。
損益を減らすため、債務不履行となる可能性が高いユーザーかどうかを借入れ前に判断を行い、
上限金の制限などによる対策実施を検討しています。
そこで、保有する過去数年分のクレジットカード利用履歴などのデータから債務不履行と
なるかどうか判別を行うAIシステムを作成します。
本AIシステムは、学習/実行に使用するデータとして、実際にカード所有者がクレジットカードで
行った取引データ（トランザクションデータ）を用います。

使用する取引データは下記（表 1）のように、28の要素からなります。
本AIモデルは、このデータセットの中から、不正なクレジットカード取引を検出する事を目的としています。

（表 1）使用するデータの属性

No	説明	補足
項目1	指定されたクレジットの金額	個人の消費者クレジットとその家族（補足）クレジットの両方が含まれる
項目2	性別	男性/女性
項目3	教育	大学院/大学/高校/その他
項目4	婚姻状況	既婚/独身/離婚/その他
項目5	年齢	年
項目6～11	過去6か月(x月～x-5月)の返済状況	未利用/全額支払済/リボ払い使用/1か月間の支払遅延/2か月間の支払遅延/・・・/8か月間の支払遅延/か月以上の支払遅延
項目12～17	過去6か月(x月～x-5月)の請求書金額	クレジットカード利用額（請求額）
項目18～23	過去6か月(x月～x-5月)の支払金額	上の項目に対し、実際に支払った金額

<処理の流れ>

本AIモデルでは、下記のような流れで、機械学習をおこなった上でデータの中から不正なトランザクションを検出します。

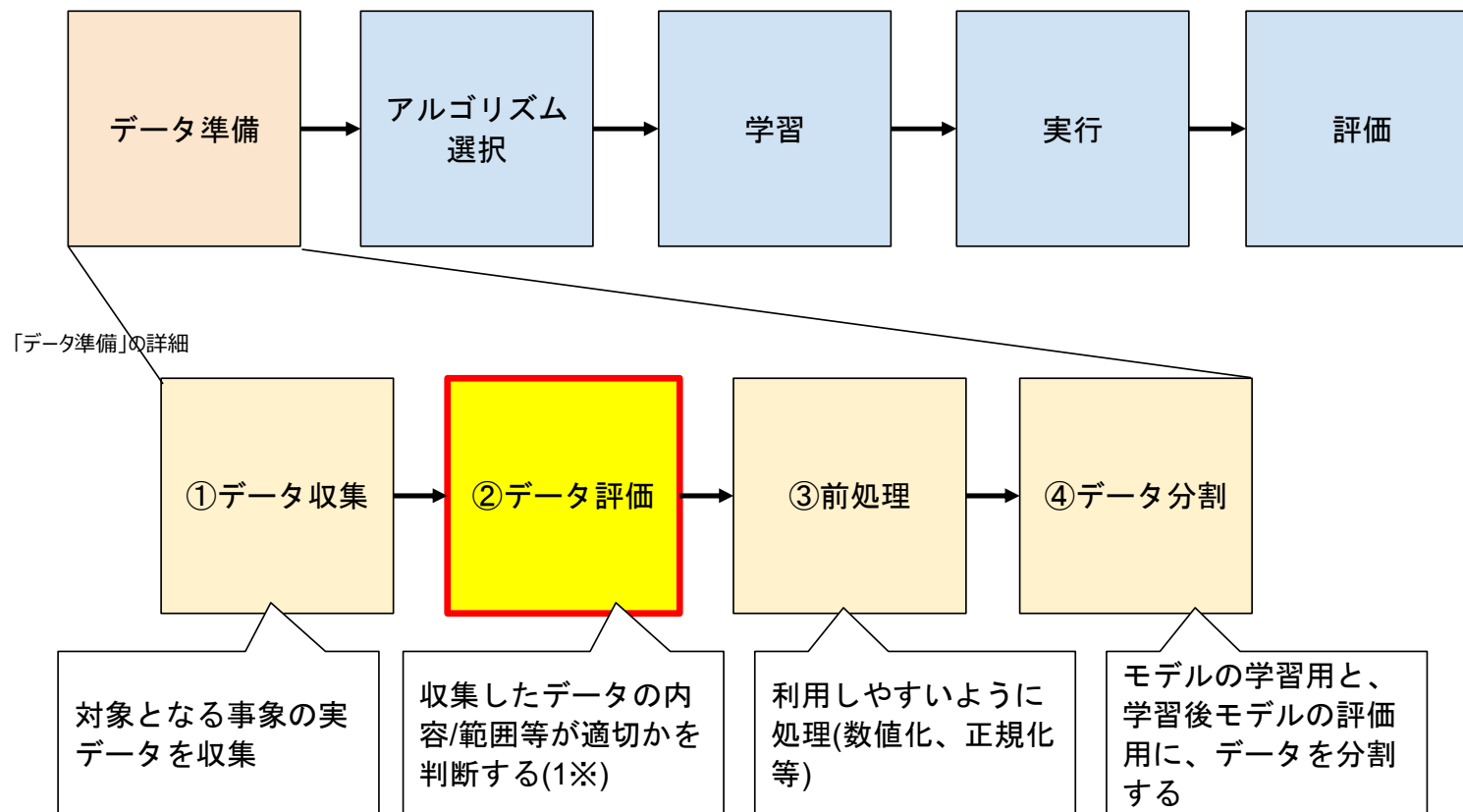
基本的な流れ

データ準備	モデルを学習/評価するためのデータを用意します
アルゴリズム選択	様々な機械学習アルゴリズムの中から、今回適用するものを選択します
学習	選択したアルゴリズムによるモデルを、学習用データを使って学習させます
実行	学習させたAIモデルに対して、実行用データを用いて処理をおこない、不正なトランザクションを検出します
評価	結果を評価する。また各種モデル同士の比較評価もおこないます

上記「データ準備」の作業詳細

データ収集	学習・実行に使うためのデータを集めます（実際のデータ）
データ評価	データに過不足がないか評価します
前処理	データを学習しやすい形に整形します
データ分割	データを学習用と評価用に分割します

基本的な流れ



（※ 1）データ評価について

＜機械学習モデルとデータ＞

データ主導となる機械学習（教師あり学習による実装）の場合、機械学習の品質を確保するためには、データに一定の品質を確保することがほぼ必須の手段となる。

そのため、学習用のデータそのものが売買や開発役割分担の対象となることがあり、このような場合には「データの品質」を単体の性質として議論する必要がある。

また不正データの意図的混入による学習結果の汚染のセキュリティリスクも指摘されており、データの出所についても何らかの広い視点からの担保が必要である。

＜データの評価方法＞

使用するデータが機械学習モデルの品質を確保出来るものであるか、の評価が機械学習モデル構築時には必要となるが、機械学習（AI）分野ではテストデータ品質の評価方法について標準的なガイドラインがまだ模索中の段階である。今回、国立研究開発法人産業技術総合研究所（産総研）が2020年6月に発表した、AIシステムの設計開発における品質マネジメントについて体系的にまとめた「機械学習品質マネジメントガイドライン」をもとに、データ品質の評価を実施し検証をおこなうものとする。

＜産総研ガイドライン＞

<https://www.cpsec.aist.go.jp/achievements/aiqm/>

本ガイドラインの品質管理の基本的な考え方

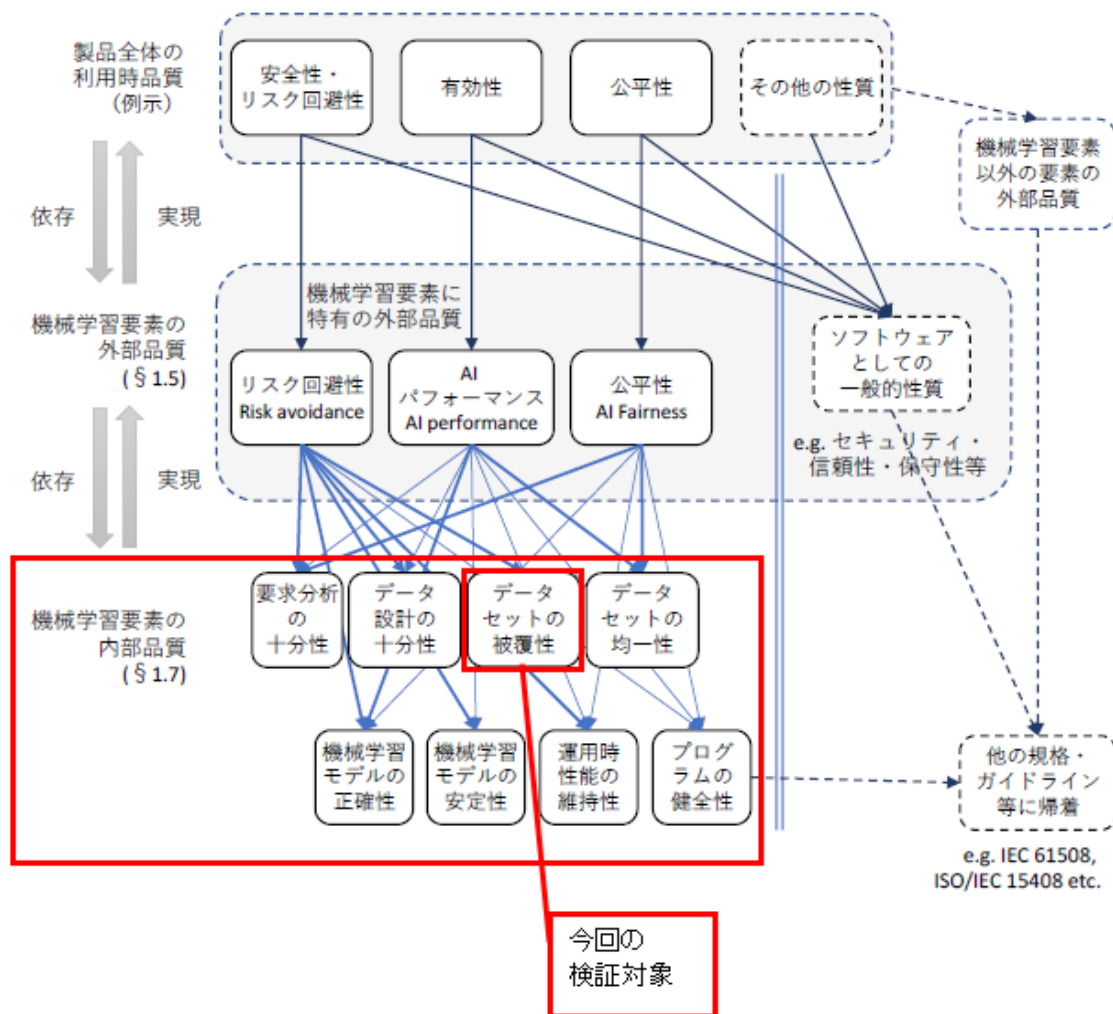
- ・「利用時品質」・・・システムがその全体として利用時に満たすことが期待される品質
- ・「外部品質」「内部品質」・・・SystemsEngineering の考え方から実際にシステムが提供する品質

「外部品質」・・・システムのうち機械学習で構築された構成要素が満たすことが期待される品質

（例）セキュリティ、信頼性、一貫性 ：定性的や定量的なものを含む

「内部品質」・・・具体的に測定したり、設計などの開発行為を通じて評価したりする品質

(例) 機械学習モデルの品質（正確性、安定性、維持性等）、データの品質（十分性、被覆性、均一性等）、など



今回のアンケートでは、内部品質のうち「データセットの被覆性」を対象とする。

<データセットの被覆性とは>

「対応すべき状況の組み合わせ」の各々に対して、状況の抜け漏れがなく、十分な量の学習データが与えられていること。

(例) 自動運転のAIモデルの場合、学習データ項目に「走行時の天気」を設ける場合、

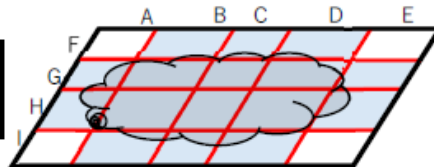
晴れ/曇り/雨/雪/霧/... の様に、実際に走行する場合に起こりうる状況をすべて網羅しているか。

雨の状況を学習データに準備していない場合、実際に雨の際に走行すると予想できない挙動を示す事が考えられる



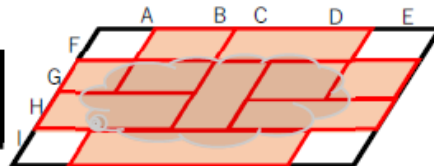
データの設計

性質1:
要求分析の
十分性



- 機械学習要素の対応する動作範囲を規定する
- 想定の入力データの範囲を特定する
 - データラベルなどの具体的な形で記述する
 - 対応しない状況とレアな状況等を峻別する

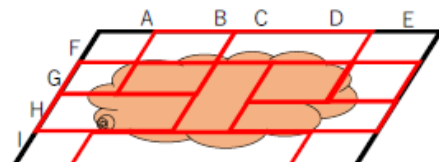
性質2:
データ設計
の十分性



- 品質管理に用いるデータ整理の細分の枠組を設計する
- 高リスクな状況の組み合わせなどを網羅的にカバーする
 - 全体として取り扱える範囲の数に設計する

データの品質

性質3:
データセット
の被覆性



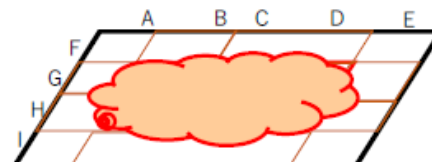
それぞれの細分した領域ごとに十分なデータが含まれることを確認する

- データの量が十分であること
- データに偏りがないこと

→ 十分にリスクなどに対応した学習訓練がされることを担保する

バランスを取って
両立させる

性質4:
データセット
の均一性



データ全体として偏り無く均一にデータが含まれることを確認する

→ モデルの全体性能を向上させる

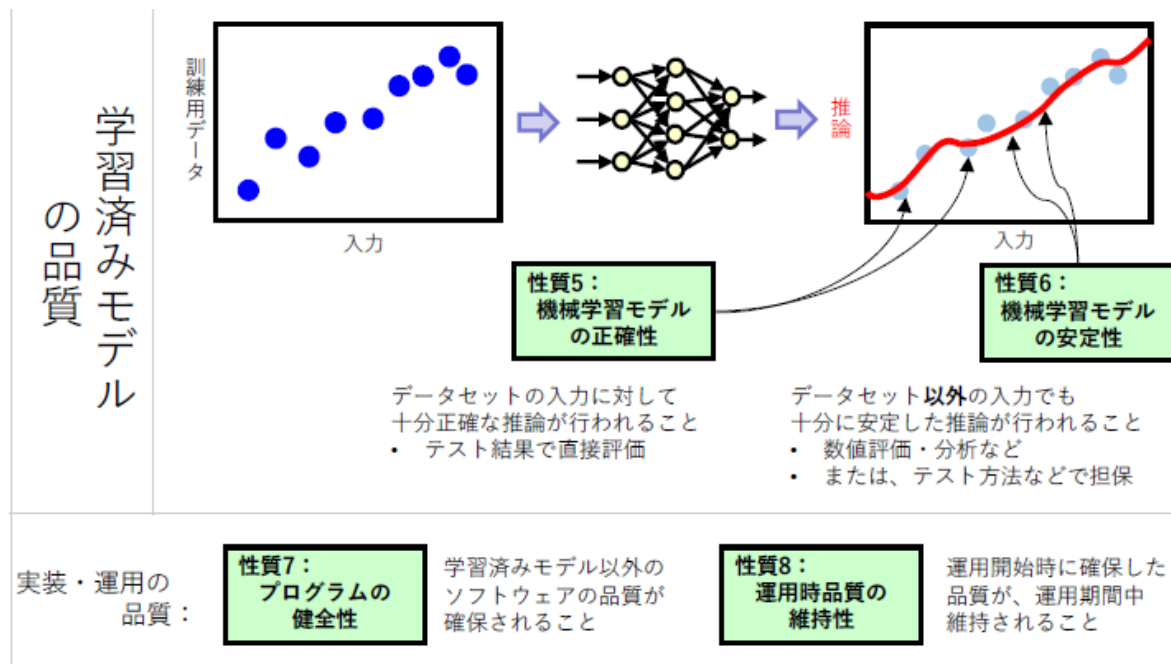


図 3: 着目する内部品質特性