生成 AI を活用したソフトウェア開発プロセスのセルフアセスメントアシスタント(AI-ProSaA)の提案

Proposal of Generative AI-assisted Process Self-assessment Assistant (AI-ProSaA)

株式会社デンソークリエイト
DENSO CREATE INC.
池永 直樹
Naoki Ikenaga

Abstract To understand the performance of the software development process and identify improvement points, there is a method where software engineers perform a self-assessment using a checklist. There is a problem with self-assessment in that the result heavily depends on the practitioner's process knowledge. Generative AI with strong natural language processing capabilities may also be effective in certain aspects of the process. In this research, the author developed “Generative AI-assisted Process Self-assessment Assistant (AI-ProSaA)” to support practitioners who lack sufficient knowledge of the process. As a result of experiments conducted in actual projects, it was confirmed that generative AI was able to provide advice at the level of an associate (with a certain level of experience) assessor, demonstrating its effectiveness in supplementing the practitioner's process knowledge.

1. はじめに

筆者はプロセス改善活動やプロセスアセスメントを担当している．プロセス改善活動を活性化させるために，開発現場の技術者自らが開発プロセスを自己診断形式で実施するプロセスアセスメント(セルフアセスメント)の展開を推進している．プロセスアセスメントの実施にはプロセスに関する十分な知識が必要であり,その有無がセルフアセスメントの成否のカギである.しかし，開発現場の技術者はそのドメイン知識には精通しているが，プロセス知識は不足していることが多い．そこで，プロセス知識の補完に生成 AI を活用することを考え，生成 AI を活用したプロセスのセルフアセスメントを研究テーマとした.

以降 2 章で研究の背景を説明し，3 章では仮説立案から生成 AI を活用したプロセスのセルフアセスメントアシスタント(AI-ProSaA)を提案する．4 章で実験方法と結果を示し，5 章では実験結果を分析し提案アプローチの効果を考察する．最後に 6 章で研究成果のまとめと今後の展望を述べる.

2. 研究の背景

2.1 セルフアセスメントの必要性

開発現場のソフトウェア開発プロセスの状態や改善点など，現状把握の有効な手段の一つとしてプロセスアセスメントがある[2].プロセスアセスメントは,プロセスアセスメントモデル(以降,PAM)を用いて自組織/プロジェクトの仕事のやり方に,改善すべき点があるか,伸ばす点があるか,組織/プロジェクト目標の達成に対してどういう状態にあるかを診断する[3].繰り返し実施することによって，仕事のやり方の問題点の発見やプロセス改善の成果の確認に役立つ[3]．このプロセスアセスメントは，アセッサーと呼ばれるプロセスの専門家によって実施される．国内では，IPA/SEC にプロセス改善部会が組織され，プロセスアセスメントとともにプロセス改善活動の普

株式会社デンソークリエイト 生産革新部
DENSO CREATE INC. Software Production Innovation Div.

名古屋市中区錦 2 丁目 14 番 19 号 名古屋伏見 K スクエア Tel:052-728-0771
2-14-19 Nishiki Naka-ku, Nagoya, Aichi, Japan

【キーワード】プロセスアセスメント、アセッサー、生成 AI、プロセス改善

及促進が図られた[4]．また，筆者が従事する自動車業界では，Automotive SPICE が業界標準の PAM として存在し，プロセスアセスメントがグローバルで年間約 1500 回(2022 年)実施されていると報告されている[5]．

プロセスアセスメントの実施には，十分なインタビュー時間が必要である．しかし，アセッサーの人数は限られているため，組織内の多数のプロジェクトを網羅的に，かつ頻度高く実施することは現実的ではない．従って，セルフアセスメントがアセッサーによるプロセスアセスメントを補完する役割として期待できる．

2.2 セルフアセスメント実施における問題

問題は，開発現場の技術者によるセルフアセスメントの結果が，開発プロセスの実態を表したものにならないことである．なぜなら，PAM はそもそも抽象度が高く，PAM を基に作ったプロセスのチェックリストを用いても，技術者のプロセスに関する知識不足を十分補完できず，チェック項目の意味を誤解釈して又は達成度合いを正しく判断せず自己診断してしまうためである．

2.3 セルフアセスメント展開時の課題

2.2 節の問題を解決し組織のプロセス改善を推進させるために，セルフアセスメント展開時の課題を以下と設定する．

課題：セルフアセスメント実施者のプロセスの知識レベルに寄らず，開発プロセスの実態を表した診断結果を得られる

3. 解決策の提案

3.1 仮説

生成 AI は要件定義，議事録管理，プログラム開発(コード生成，ペアプログラミングなど)，レビュー，テストなどソフトウェアエンジニアリングへの適用が進んでいる[6]．

その状況のもとプロセスアセスメントにおいても，生成 AI がアセッサーを支援するアセスメントツール[7]などが登場し始めている．そこで，これらの事例のように，生成 AI の特徴である自然言語処理，知識の広さ，文章読解力，コンテンツ生成は，自然言語中心で扱われるセルフアセスメントの精度向上に寄与する可能性が高いのではないかと考えた．よって，本研究の仮説を以下と設定する．

仮説：セルフアセスメント実施者のプロセス知識が不十分でも，生成 AI のサポートがあれば自己診断の精度が向上するのではないか？

3.2 生成 AI を活用したセルフアセスメントアシスタント(AI-ProSaA)の提案

セルフアセスメントへの生成 AI の活用方法として，最終的な達成度の判断はセルフアセスメント実施者に任せ，チェック項目に対する「評定の根拠」に対して生成 AI にアドバイスさせるセルフアセスメントを支援する方式を考えた．そこで，「生成 AI を活用するプロセス」「チェック項目の構成」「プロンプトテンプレート」をセットとした AI-ProSaA(Generative AI-assisted Process Self-assessment Assistant)を提案する．この方式を採用する理由を以下に述べる．

セキュリティ監査業務に対する生成 AI の性能を評価した先行研究がある[8]．この研究では，成果物そのものを生成 AI に入力し適合／不適合を回答させるものであったが，監査性能は十分ではなかったと報告されている．また，プロセスアセスメントは成果物のみを基にした単純な OK／NG の判断でなく，アセスメント目的やプロジェクトコンテキストを考慮した上で出来栄えを診断する必要があり，より複合的な判断が求められる．さらに，セルフアセスメントを通じて開発現場のプロセスに対する意識や知識を向上させることも重要である．

よって，生成 AI に最終的な達成度の判断を行わせるのではなく，開発現場によるセルフアセスメントを支援することに活用すべきと考えた．

(1) 生成 AI を活用するプロセス

AI-ProSaA を適用したセルフアセスメントのプロセスを図 1 に示す．本プロセスでは，生成 AI の活用箇所をチェック結果記入時のタイミングとし，ここで「評定の根拠」に対するアドバイスを出力させる．

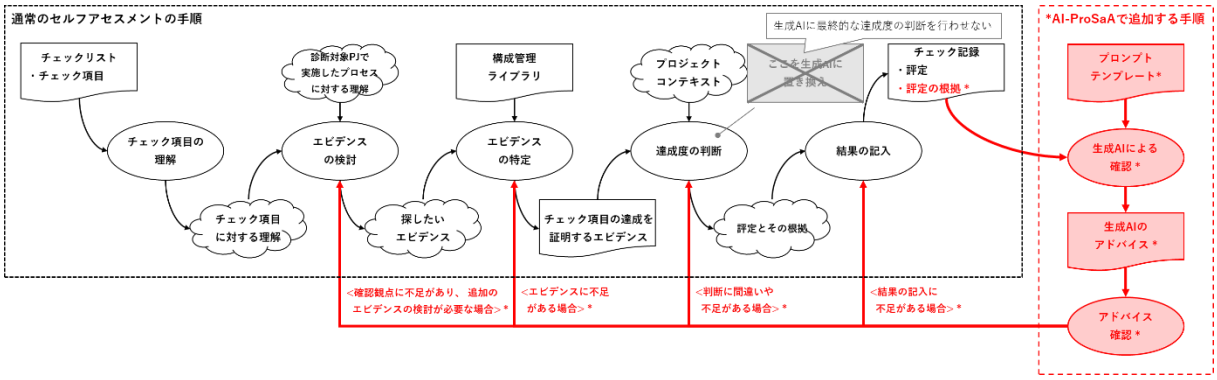

図 1 「AI-ProSaA」を適用したセルフアセスメントのプロセス

(2) チェック項目の構成

チェック項目は「設問」「設問の補足」「評定」「評定の根拠」で構成する．「設問の補足」は，自己診断時にセルフアセスメント実施者のプロセスの知識を補う目的で用意する．「評定」と「評定の根拠」をセルフアセスメント実施者に回答してもらう．

ID	設問	設問の補足	評定	評定の根拠
MAN.3-1	プロジェクトで達成すべき目標を利害関係者と合意できていますか？	利害関係者には、プロジェクトメンバー、上級管理者、顧客、委託先、開発環境のサポート窓口、関連する他のプロジェクトなどが考えられる。	F/L/P/Nのいずれかを選択する	F/L/P/Nを選択した根拠を記入する

図 2 チェック項目の例

(3) セルフアセスメントで生成 AI を有効活用するためのプロンプトテンプレート

生成 AI は自己診断結果に機密情報が含まれている可能性があることから，筆者の所属組織で利用が認められている GPT-4o の使用を前提とする．図 3 はプロンプトの全体像である．

プロンプト	
あなたはソフトウェア開発プロセス（Automotive SPICE、ISO/IEC12207、CMMIなど）の専門家です。	①ペルソナ設定
プロジェクト活動について、プロジェクトメンバー自身がチェックリストを用いてソフトウェア開発プロセスを自己診断しています。 この自己診断は、プロセスの定義状況を問うものではなく、プロジェクトでのプロセスの実施を問うものです。	②コンテキスト提供
[設問] に続く問いについて、[評定の根拠] に続く回答を得ました。得た回答に対して、プロンプトで指示するStep1,2,3,4の手順を実施して下さい。	
【Step1】 ノイズとなる情報を除去するために、[評定の根拠] から [設問の補足] も考慮して、[設問] に無関係な情報を抽出し、箇条書きで回答して下さい。 ＃除外した[設問]に無関係な情報 ・xxx：yyy（xxxは除外した情報、yyyはその理由。ない場合は「なし」のみ記載）	③無関係情報の抽出
【Step2】 「#要約の考慮点」を考慮して、[評定の根拠] を要約してください。 #要約の考慮点 ・Step1の「＃除外した[設問]に無関係な情報」を要約に含めない。 ：（省略） ＃要約 xxx（400字以内）	④「評定の根拠」の要約
【Step3】 Step2の要約結果を元に、[設問の補足] および「#達成度評価の考慮点」を考慮して、達成度を評価してください。 達成度は、F（十分達成）、L（おおむね達成）、P（部分的に達成）、N（未達成）、NR（評定不能）で回答して下さい。 FLPNは、ISO/IEC33020のプロセス属性評定の尺度のことです。NRは、要約結果の情報では達成度を評価できない場合に選択してください。 また、達成度の判断理由も回答してください。 #達成度評価の考慮点 ・Step2の要約結果がプロセス定義に関する内容のみの場合、NRと評定する。 ：（省略） ＃達成度 F/L/P/N/NR のいずれか。 ＃達成度の判断理由 xxx（400字以内）	⑤要約を基に評定（その根拠を含む）
【Step4】 Step3の達成度および判断理由をもとに、「#アドバイスの考慮点」を満たすように、正しい評定に向けたアドバイスをしてください。 #アドバイスの考慮点 ・Step3の達成度がNRの場合、NRと判断した理由のみを出力して下さい（この場合、箇条は一つ）。 ：（省略） ＃アドバイス 1.xxx（優先度が高い順に最大5個まで。各箇条は100字以内）	⑥評定結果を基にアドバイスを生成
[設問] *<設問の内容>* [設問の補足] *<設問の補足の内容>* [評定の根拠] *<評定の根拠の内容>*	【入力】

図 3 プロンプトテンプレート

生成 AI の活用においてプロンプトパターンは数多く考案されており，Prompt engineering[10]の戦術，Prompt Engineering Guide[11]のテクニック，プロンプトパターンカタログ[12]などがある．筆者はこれらから，ペルソナ設定，コンテキスト提供，出力テンプレート，Chain of Thought などを組み合わせてプロンプトテンプレートを作成した．

以下に本プロンプトテンプレートでポイントと考える内容について説明する．

【専門知識】

汎用 AI から有用なアドバイスを得るには，プロセス知識やその組織や製品分野におけるプロセスの特徴などの専門知識が必要である．チェック項目の構成要素の一つである「設問の補足」がそれに該当するため，これを生成 AI に入力する．そして，「設問の補足」から引用した形でアドバイスが出力されるようにする．

【アセッサーの思考ステップ】

初期のプロンプトは要約させた後にアドバイス生成の流れとしていた．しかし，指示をいくら調整してもアドバイスの質が向上しなかった．そこで，アセッサーの思考フローに基づき推論ステップを分割し，要約を基に評定するステップ(図 3 の⑤)を追加した．

【多層的アプローチ】

セルフアセスメント実施者のプロセスに関する知識が不足している場合，図 4 のように設問に関係のない情報が「評定の根拠」に記述される恐れがある．特に“設問範囲内であるがプロセス能力レベル 3 に該当する情報”が記述された際にこれを不用な情報として扱わせることが困難であった．これらは，コンテキスト提供や一つの検出の仕掛けだけではすり抜けが発生していた．そこで，十分なコンテキストを提供(図 3 の②)した上で，無関係な情報を抽出するステップ(図 3 の③)，評定ステップ(図 3 の⑤)に「プロセス定義に関する記述だけなら NR(Not Rated)と判断する指示」など検出のための複数の仕掛けを追加した．このようにして，いずれかのステップで不用な情報として扱わせるようにした．

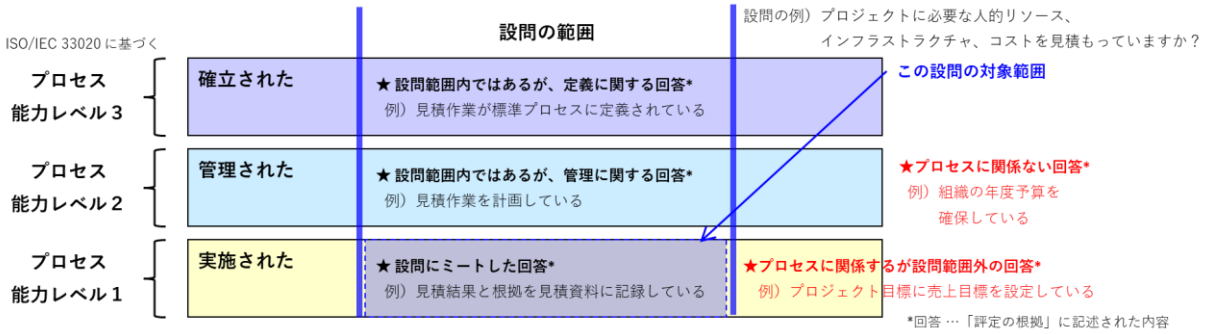

図 4 セルフアセスメントの「設問」と「評定の根拠」に記述された内容との関係

3.3 研究課題

3.1 の仮説の実現案である「AI-ProSaA」の有効性を評価するための研究課題(RQ: Research Question)を以下と設定する．

RQ1:アセッサーと同じレベル(見逃し・不要な指摘ゼロ)でアドバイスできるか

RQ2:セルフアセスメント実施者にとって納得感のあるアドバイスを 2/3 以上提供できるか

納得感の有無を比べたときに明らかに上回っていることを確認するために，地方自治体などで重要事項を決議する場合に採用される「特別多数決」の 2/3 を基準値とする．

RQ3:生成 AI のアドバイスを受けて診断結果が 1/3 以上見直されるか

診断結果の見直しは RQ2 で設定した納得感より少ないと想定される．なぜなら，アドバイスにより「何らかの気づきはあるが評定が変わる程ではない」，「評定の根拠に記入しなかったことに対してアドバイスを受けたので評定は変わらない」と考える場合があるためである．そこで，納得感を得たアドバイスのうち半数以上が見直しされれば有効なアドバイスを出力できたと考え，本基準値を RQ2 の半分の 1/3 とする．

4. 実験

4.1 実験方法

セルフアセスメントを実施する立場であるプロジェクトマネージャー(又はリーダー)を対象に実験を行う．チェックリストを用いて，実際のプロジェクトを自己診断してもらう．それに対して生成 AI が出力したアドバイスをアセッサー及びセルフアセスメント実施者が評価する．ここで，アセッサーは Automotive SPICE Competent Assessor 資格[1]を保有する筆者が担当する．

4.1.1 評価観点

本実験における 3.3 節で述べた RQ の評価観点を示す．

(1) RQ1 の評価観点

生成 AI のアドバイスに見逃しや不要な指摘がないかをアセッサーが評価する．

- アドバイスの内容に見逃しがあったチェック項目
- アドバイスに不要な指摘が含まれているチェック項目

(2) RQ2 の評価観点

生成 AI のアドバイスをセルフアセスメント実施者が利用することから，ISO25010 の利用時品質の有効性/効率性/満足性を用いて評価観点を定義し，セルフアセスメント実施者が評価する．

表 1 利用時品質を用いた評価観点

利用時品質の特性	評価観点
有効性	理解できたか
効率性	分量は適量であったか
満足性	自己診断を見直すキッカケになったか

(3) RQ3 の評価観点

生成 AI のアドバイスを受けて自己診断結果が見直されたかを確認する．

4.1.2 実験対象のプロセスとチェック項目

実験対象とするプロセスは Automotive SPICE PAM V4.0 から，管理プロセス群の MAN.3 プロジェクト管理，支援プロセス群の SUP.8 構成管理，エンジニアリングプロセス群の SWE.1 ソフトウェア要求分析を選択する．プロセスの特性により傾向が偏るリスクを低減するために，3 つのプロセス群から 1 プロセスずつ抽出した．能力レベル 1 を対象とする．

「設問」は，Automotive SPICE 4.0 実践ガイドブック[9]に掲載されているチェック項目を使用する．各プロセスのチェック項目数は，MAN.3：12 項目，SUP.8：12 項目，SWE.1：9 項目である．また「設問の補足」は実践ガイドブックの解説文に筆者が加筆修正を加えたものである．

4.1.3 実験手順

次の手順で実験を実施する．

① セルフアセスメント実施者がチェックリストに回答する
② アセッサーが「評定の根拠」を確認し，アドバイス(期待値)を作成する
③ 「設問」「設問の補足」「評定の根拠」を生成 AI に入力し，アドバイスを出力させる
④ セルフアセスメント実施者が生成 AI のアドバイスを受けて自己診断結果を見直す
⑤ アセッサーが手順②の期待値をもとに生成 AI のアドバイスを評価する
⑥ セルフアセスメント実施者が生成 AI のアドバイスを評価する

4.2 実験データ

セルフアセスメント実施者 7 名が実際のプロジェクトを対象に自己診断した回答(#1-7)，及び筆者が恣意的に作成した回答(test)の合計 161 件の実験データを得た．

表 2 セルフアセスメントの実験データ

	#1	#2	#3	#4	#5	#6	#7	test	合計
MAN.3	0	12	12	12	0	12	11	6	65
SUP.8	0	12	12	12	12	12	0	0	60
SWE.1	9	9	0	0	9	0	9	0	36
合計	9	33	24	24	21	24	20	6	161

4.3 実験結果

3.3 節で述べた RQ の実験結果を示す．なお，図 5 は生成 AI の出力例である．

```
#回答
## Step1: 除外する[設問]に無関係な情報の抽出
・なし：設問に無関係な情報は含まれていません。
## Step2: 要約
プロジェクトに必要な人的リソースとコストは「見積もり計画書」で見積もられ、インフラストラクチャは「4.開発環境」および「5.リソース計画」で見積もられている。
## Step3: 達成度評価
### 達成度
P（部分的に達成）
### 達成度の判断理由
「プロジェクトに必要な人的リソース、インフラストラクチャ、コストを見積もっていますか？」という設問に対し、人的リソースとコスト、およびインフラストラクチャの見積もりが「見積もり計画書」、「4.開発環境」、「5.リソース計画」で行われていることが確認できる。しかし、見積もりの手法や根拠についての具体的な説明がないため、客観的に説明可能な見積もりが実施されているかどうかの確認が不十分である。
## Step4: アドバイス
1. 見積もり手法やそれに基づく根拠が客観的に説明可能であることを確認してください。
2. 見積もりに使用された方法やデータの詳細を見積もり計画書等に明確に記載していることを確認してください。
```

図 5 生成 AI の出力例

(1) RQ1 の実験結果

実験手順③で得られたアドバイスに見逃しや不要な指摘がないかを計測した結果を表 3 に示す．なお，一つのチェック項目に対して複数のアドバイス項目が出力されることから，図 6 のようにアドバイス項目ごとに分類している．

表 3 見逃し・不要な指摘の計測結果

	チェック項目	アドバイス項目数
見逃し	45/161 項目(28%)	46 個
不要な指摘	50/161 項目(31%)	62 個

設問	生成AIが出力したアドバイス	アドバイスの分類
プロジェクトに必要な人的リソース、インフラストラクチャ、コストを見積もっていますか？	1. 見積もり手法やそれに基づく根拠が客観的に説明可能であることを確認してください。	適切な指摘
	2. 見積もりに使用された方法やデータの詳細を見積もり計画書等に明確に記載していることを確認してください。	不要な指摘
	【分析時メモ】管理・支援プロセスの活動の見積りについてのアドバイスが出力されなかった	見逃し

図 6 チェック項目に対するアドバイスの分類例

(2) RQ2 の実験結果

実験手順⑥のセルフアセスメント実施者が生成 AI のアドバイスを評価した結果を図 7 に示す．

- 「理解できたか(有効性)」の良い・やや良いの割合：79%
- 「分量は適切であったか(効率性)」の良い・やや良いの割合：93%
- 「自己診断を見直すキッカケになったか(満足性)」の良い・やや良いの割合：74%

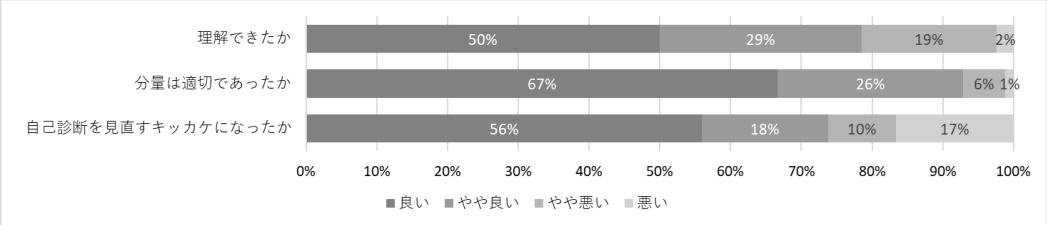

図 7 生成 AI のアドバイスに対するセルフアセスメント実施者の評価

(3) RQ3 の実験結果

実験手順④の生成 AI がアドバイスを出力したチェック項目に対して，セルフアセスメント実施者が自己診断結果を見直した割合は 32%であった．

5. 考察

5.1 RQ の考察

3.3 節で述べた RQ に対する考察を述べる．

(1) RQ1:アセッサーと同じレベル(見逃し・不要な指摘ゼロ)でアドバイスできるか の考察

実験から，見逃しが 45 項目，不要な指摘が 50 項目でありゼロではないため RQ1 に対し完全に有効だったと言えないが，初級アセッサーと同等レベルであることが確認できた.

アセッサーと比較すると見逃しは 45/161 項目(28%)，不要な指摘は 50/161 項目(31%)とそれぞれ 3 割程度多かった．初級アセッサーのデータがあり，見逃し率 28%・不要な指摘率 4%である．これと比較すると見逃しは同水準であり，不要な指摘は多い結果となっている．なお，この初級アセッサーのデータは，筆者が所属する組織の Automotive SPICE Provisional Assessor 資格[1]を保有し、プロセス改善業務経験は十分ありアセスメント経験が数回のメンバーの平均値である.

セルフアセスメントの性質上，過剰に指摘する分(不要な指摘)については許容できるが，見逃しは問題である．初級アセッサーと同等レベルであったので実務で活用可能なレベルと判断するが，精度向上の余地は十分ある．以下に精度向上に向けた分析を示す.

■見逃し

いずれのチェック項目でも同じ内容を見逃す傾向があることが確認できた．見逃した内容はいずれも「設問の補足」に記述されている内容であった．今回「設問の補足」はセルフアセスメント実施者が設問に回答する際にプロセスの理解の助けとなるように，解説書のように読み物として記述した．生成 AI に「設問の補足」をそのまま入力していたが，診断観点リストとして要点に絞って入力すると見逃しが軽減できるのではないかと考える．本実験データでこの対策が機能した場合，アドバイス項目数 46→10 個・チェック項目 45→9/161 項目(28→6%)になる試算である.

■不要な指摘

不要な指摘は 4 つに分類できたので表 4 にその分類結果を示す.

表 4 不要な指摘の分類結果

#	分類	説明	アドバイス項目数
1	オウム返し	「評定の根拠」に“○○ができていない”と記述している内容をそのまま“○○を確認して下さい”と出力	28/62 個(45%)
2	過剰な要求	その設問の範囲内であるが過剰と考えられる内容	25/62 個(40%)
3	無関係	その設問の範囲外で明らかに無関係と判断できる内容	2/62 個(3%)
4	見当違い	その設問の範囲内であるが「評定の根拠」の内容に対して明らかに見当違いと判断できる内容	7/62 個(12%)

分類#1 はプロンプトチューニングで対策できると考えられる．本実験データではこの対策により，分類#1 の不要なアドバイス項目 28 個が排除されるため，アドバイス項目数 62→34 個・チェック項目 50→29/161 項目(31→18%)になる試算である.

他方，分類#2-4 については対策が困難であると考えられる．分類#2 はその設問の範囲内であるためアドバイスとして出力されても問題ではない．分類#3,4 はアドバイスの質を落とすものであるが，セルフアセスメント実施者のプロセス知識に関わらず明らかに不要な内容と判断できると考えられることから実害はないと判断する.

(2) RQ2:セルフアセスメント実施者にとって納得感のあるアドバイスを 2/3 以上提供できるかの考察

実験から,「やや良い」以上の評価が有効性:79%, 効率性:93%, 満足性:74%であり，基準値の 2/3 をいずれも上回っていることから，RQ2 に対し有効だったと判断する.「知識が得られた」「間違いに気づいた」とのコメントも得られたことから，納得感があったものと判断できる.

一方で,「できていないと記述していることに対して再度確認せよとアドバイスされる」「アドバイスが曖昧」などのコメントもあり，これらが評価を下げる要因となっている．前者については(1)RQ1 の不要な指摘の考察で述べた分類#1 の対策を講じることで対策が可能である．後者については生成 AI のアドバイスで解決するのではなく，「設問の補足」を活用するなどして対策するのが良いと考える.

(3) RQ3:生成 AI のアドバイスを受けて診断結果が 1/3 以上見直されるか の考察

実験から，実績が 32%であり，基準値の 1/3 に対して僅か 1%の差異(近似値)と捉え，RQ3 に対

し概ね有効だったと判断する.

見直し結果はチェック項目に対して適切な評価になり，実施された開発プロセスの実態を表した結果に修正されていることが確認できた.

セルフアセスメント実施者からは,「気づきを得た」「質問に回答できないことが分かった」などのコメントが得られた. また，想定通り「メモとして最小限しか書いておらず，書いていないことについてアドバイスが出力されていた」というコメントも確認できた.

5.2 妥当性への脅威

実験データは主に筆者所属組織の要員のものであり，自組織では有効であるものの，仮に他組織で実験した場合は，実験データとアセスメント実施者が異なるため，実験結果が変わる恐れがある. 今後，セルフアセスメント実施者を増やしより多くの実験データでの傾向を確認したい.

また，他プロセスや能力レベルについても実験することで，プロセスや能力レベルによる傾向の違いがないことも確認したい.

6. まとめ

6.1 研究成果

ソフトウェア技術者自らがソフトウェア開発プロセスを自己診断するセルフアセスメントを展開するにあたり，2 章で述べた「課題 : セルフアセスメント実施者のプロセスの知識レベルに寄らず，開発プロセスの実態を表した診断結果を得られる」を解決する必要がある. そこで本研究では, セルフアセスメントの実施を支援する目的での生成 AI の活用としてアシスタント方式である AI-ProSaA を考え，その有効性を評価した. 実験の結果，初級アセッサーと同等レベルで納得感のあるアドバイスの出力が可能であり，自己診断の見直しにも役立つことが確認できた. 従って，妥当性への脅威はあるが，RQ1～3 の結果により AI-ProSaA の有効性が確認できたことで仮説は支持され，生成 AI にアドバイスさせる支援方式が，セルフアセスメント実施における問題を解決できる(自己診断結果が開発プロセスの実態を表したものに近づく)と判断する.

6.2 今後の展望

セルフアセスメント実施の支援に生成 AI が活用できることが確認できたが, 初級アセッサーレベルになっており改善の余地が残っている. 以下のように取り組み，生成 AI によるアドバイスの質をアセッサー(筆者)のレベルまで向上させる.

(1) 5.1 節で述べた見逃し・不要な指摘に対する対策の実施
(2) 新しい生成 AI モデルの活用

参考文献

[1] international Assessor Certification Scheme, https://intacs.info/
[2] IPA/SEC, プロセス改善ナビゲーションガイド ～虎の巻編～, 2009/2/25
[3] IPA/SEC, プロセス改善ナビゲーションガイド ～プロセス診断活用編～, 2007/3/30
[4] 堀田勝美,日本におけるプロセスアセスメント活動, 情報処理学会 短期集中セミナー, 2020/1/10
[5] Jan Morenzin, Automotive SPICE® News and data from VDA QMC, 1st Asia SPICE Conference
[6] AI を用いたソフトウェア開発, https://www.ipa.go.jp/digital/ai/software-engineering.html
[7] Assessor Academy, AXIOM -次世代型アセスメントツール-, https://assessor.co.jp/axiom/
[8] 多田麻沙子,徳本晋,栗田太郎,石川冬樹, ISO27017 に基づくクラウドセキュリティ監査業務に対する LLM の性能, ソフトウェア・シンポジウム 2024
[9] Business Cube & Partners, Automotive SPICE 4.0 実践ガイドブック 入門編, 日経 BP, 2024/1/22
[10] OpenAI prompt engineering, https://platform.openai.com/docs/guides/prompt-engineering
[11] Prompt Engineering Guide, https://www.promptingguide.ai/
[12] Jules White et al., "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT", 2023.