

生成AIを活用したテスト分析の実験的評価と プロンプト設計の考察

2025年度ソフトウェア品質管理研究会 研究コース5「人工知能とソフトウェア品質」 テスト分析チーム

研究員：川合 泰史（富士電機株式会社）
齋藤 美穂（株式会社ベリサーブ）
倉田 優輝（株式会社日立ソリューションズ・クリエイト）
主査：石川 冬樹（国立情報学研究所）
副主査：徳本 晋（富士通株式会社）
アドバイザー：栗田 太郎（フリー株式会社）

背景

ソフトウェアの大規模化や複雑化により、
テスト分析の効率化と品質向上が求められている。

従来のテスト分析

- ・仕様書が大量
- ・人手/経験依存
→見落としリスク

実験

生成AI活用

- ・要件から候補抽出
- ・分析のヒント提示
- ・LLMによる支援



課題

テスト分析には人材・知識不足、時間と労力、網羅と効率の両立が難しいという課題がある。

① 人材・知識不足

- ・ 属人化
- ・ 育成コスト増大
- ・ 品質のばらつき



② 時間と労力

- ・ 手作業中心
- ・ リリース短期化
- ・ 工数逼迫



③ 網羅と効率の両立

- ・ 網羅率向上が困難
- ・ 優先順位付け難
- ・ 変更追従が困難



⇒ テスト分析の高度化と効率化の両立がボトルネック

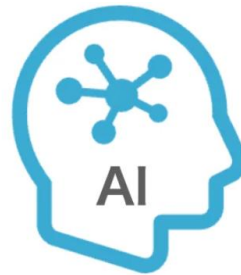
目的

本研究では生成AIを活用したテスト分析の手法を提案し、その有効性を評価する。

要求仕様書

生成AIによる
テスト分析

人材テスト分析
との比較



- ① テスト分析結果の品質評価（テスト対象、目的条件、優先度、カバレッジ）
- ② 作業時間の比較（人材 vs 生成AI）

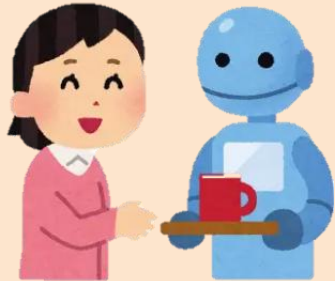
⇒ 生成AIによるテスト分析の実用化可能性を検証する

仮説

生成AIはテスト分析の効率化と品質向上に寄与するが、プロンプト設計に依存する。

生成AIの能力

- ・仕様書から観点抽出
- ・短時間で知見提示
- ・未経験者支援



テスト分析観点
なし

想定外の分析
品質低下
見当違い



テスト分析観点
あり

観点通り
効果的な試験
網羅的
明確化



プロンプトで
分析観点を
明確に指示



仮説 品質要求に応じた分析観点を明示することで
経験やスキルを問わない均一なテスト分析が可能

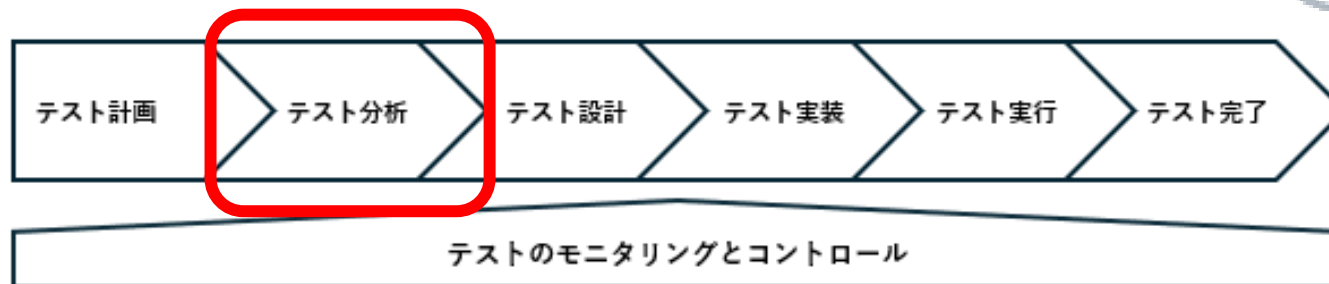
提案手法：テスト分析の定義

仮説を検証するために、まずJSTQB※1の定義する「テスト分析」について調査を実施

■ JSTQBにおける「テスト分析」の定義

- テスト対象を分析し、対象機能や特性を特定
- テスト条件を定義し、優先順位を決定
- リスク評価で重点検証領域を明確化
- カバレッジ基準でテスト範囲を明確化
- 「何をテストすべきか」を定義する

そもそも
テスト分析って
なんだろう？



※1 JSTQB

国際的なソフトウェアテスト技術者資格認定組織「ISTQB」の日本加盟団体（Japan Software Testing Qualifications Board）が運営する、世界基準のソフトウェアテスト技術者資格

提案手法：テスト分析観点

JSTQBが定義するテスト分析の実施内容を参考に、
テスト分析の観点を決定

ポイント

自分たちの業務で使える
実用的なテスト分析



プロンプト
テンプレート
(分析用/採点用)

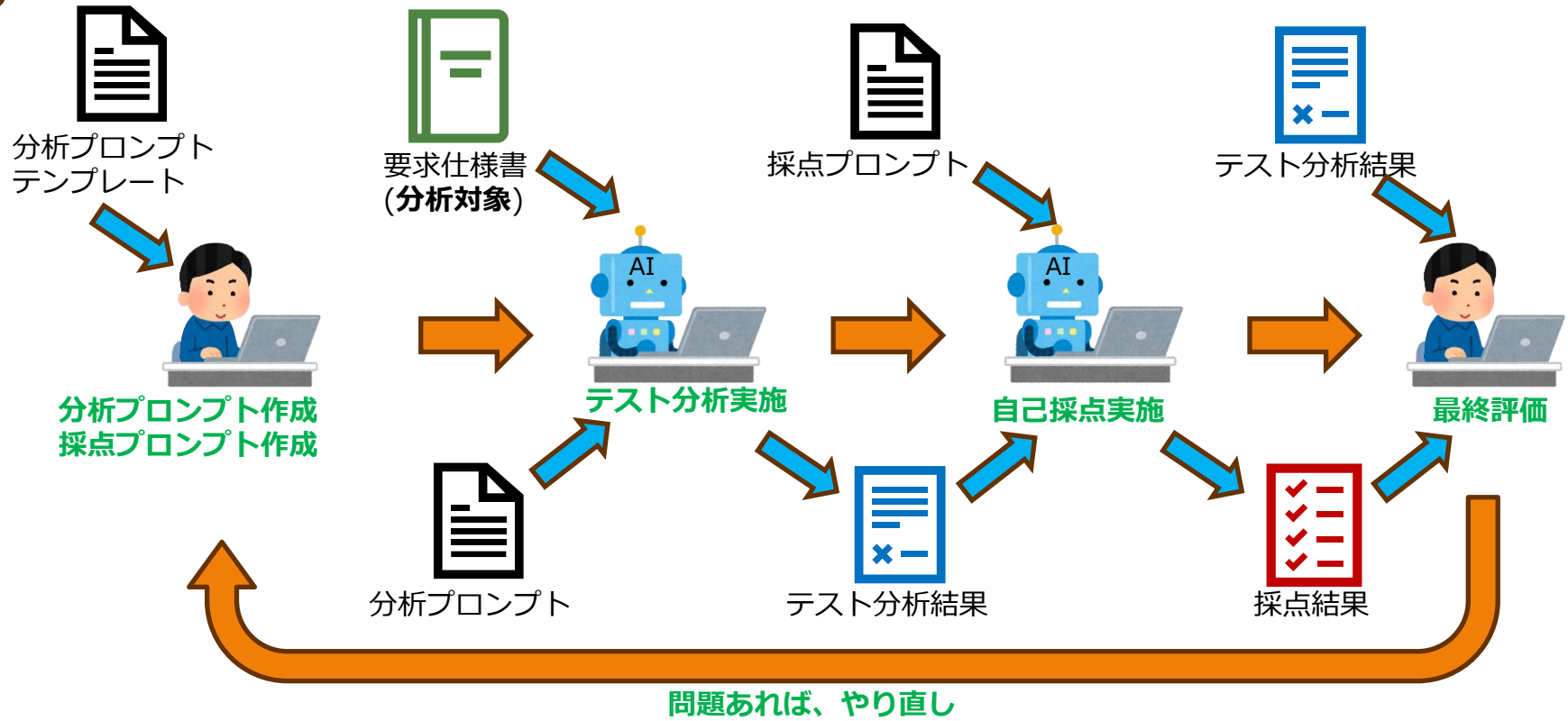


テスト分析観点

No.	項目	説明
1	テストの対象	テスト範囲の明確化と分析漏れの防止
2	テストの目的	テストの方向性の決定（機能、安全性、他）
3	テスト条件	入力値や環境を定義して再現性と信頼性を担保
4	優先度	不具合の影響度を評価して優先順位を決定
5	トレーサビリティ・カバーレッジ	要求仕様との対応を明示し網羅性を確保

提案手法：テスト分析プロセス

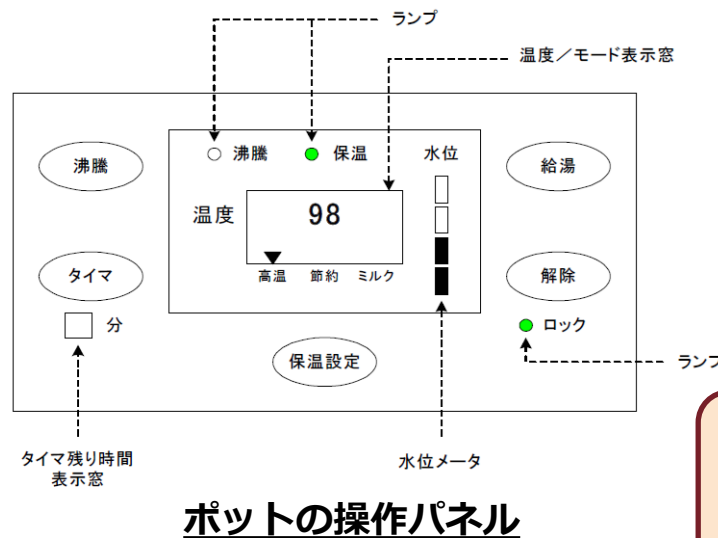
AIがテスト分析を実施して結果を自己採点し、人間が最終的に評価するテスト分析プロセスを提案



テスト分析と結果の評価をAIで自動化

実験方法：分析対象

実験に使用する分析対象としてAIが一回の思考で分析を完了できる分量の要求仕様書を選定



操作パネルでポットの操作や状態の確認が可能。

No.	項目	説明
1	ページ数	14ページ
2	機能数	12機能 (沸騰、保温、給湯、タイマなど)
3	UI	操作パネル (ボタン/LEDランプ/温度表示)

※使用したAI
Gemini 2.0 Flash Experimental
Gemini 3.0

実験方法：分析方針

提案手法の手順を用いて分析方針による結果の違いを比較

分析プロンプトテンプレート

```
#指示
テスト分析を行ってください。
#方針
(ここに分析対象毎の方針を記載)
#出力項目
・テスト対象: ~~~~~
・テスト目的: ~~~~~
・テスト条件: ~~~~~
・優先度 : ~~~~~
```

分析プロンプトテンプレートの#方針を変更して分析方針を設定

例) リスクドリブンの場合

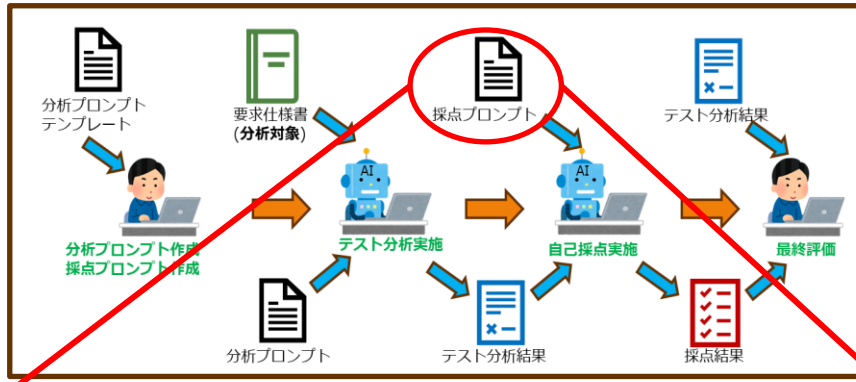
- ・リスクベースドテストの観点からテスト分析を実施。
- ・潜在的な欠陥や業務停止リスクを特定すること。
- ・リスクごとにテスト重点度・対策方針を明記すること。
- ・想定リスク一覧（原因・影響・検出可能性）を出力すること。
- ・
- ・

比較対象

No.	テスト分析方針	説明
1	指定なし(人力)	人力(経験30年のQA)によるテスト分析する
2	指定なし(AI)	分析方針を指定せずに「テスト分析を行うこと」のみを記載する
3	ユーザーストーリー	対象製品のユーザーとそのストーリーを想定し、必要なテストを分析する
4	リスクベース	リスクを想定し優先度を設定、優先度ごとにテスト内容を定義する
5	リスクドリブン	リスクを基準にテストの優先順位と範囲を決める分析を行う
6	体系的手法強調	国際規格（ISO/IEC/IEEE 29119）に準拠したテスト分析を行う
7	抽象・戦略	初期段階で決定するテスト戦略立案の観点からテスト分析を行う
8	開発時品質(主特性・副特性)	ISO/IEC 25010の品質特性に基づき、外部品質のテスト分析を行う
9	開発時品質(副特性)	開発時品質の副特性に注目し、ISO/IEC 25010に基づきテスト分析を行う
10	HAYST法	HAYST法で3階層からテスト条件を抽出し、抜け漏れを防ぐ

実験方法：採点プロンプト

■ テスト分析観点を基に、採点プロンプトを作成



あなたはソフトウェアテストの専門家AIです。
先程実行したテスト分析結果に対して、以下の観点で採点する。

#定義

- ・添付仕様書は要求仕様書であり、~~~~

#採点項目

- ①テスト対象
 - ・テストによって検証すべき~~~~
- ②テスト目的
 - ・テストの主な狙い~~~~
- ③テスト条件
- ④優先度
- ⑤トレーサビリティ/カバレッジ

#採点基準

- 5点：非常に良い~~~~~

No	項目	説明
1	定義	テスト分析の前提条件を記載 <ul style="list-style-type: none"> ・「テスト分析」の定義 ・入力ファイルの定義 ・採点方法 ・出力方法
2	採点項目	JSTQBや発表者の業務内容を元に作成した、 テスト分析で注目すべき観点 <ul style="list-style-type: none"> ・テスト対象 ・テスト目的 ・テスト条件 ・優先度 ・トレーサビリティ/カバレッジ
3	採点基準	「採点項目」の遵守度合いを採点 <ul style="list-style-type: none"> ・5点：非常に良い ～ ・1点：非常に悪い

採点プロンプト

実験結果まとめ

■ 実験方法に基づいて実験したところ…

- 人力と比較してAIによるテスト分析結果は概ね人力と同等以上の品質を達成できた
- 実行時間の大幅短縮（人力：360分 → AI：平均25分）

区分	No	テスト分析方針	テスト対象	テスト目的	テスト条件	優先度	トレーサビリティ・カバレッジ	平均	実行時間(分)
人力	1	指定なし(人力)	3.67	4.67	4.50	1.33	4.50	3.89	360
	2	指定なし(AI)	4.33	3.33	4.33	1.00	3.50	3.47	20
AI	3	ユーザーストーリー	5.00	4.67	4.50	5.00	3.25	4.42	23
	4	リスクベース	5.00	5.00	4.67	5.00	4.50	4.79	26
	5	リスクドリブン	5.00	5.00	4.50	5.00	4.75	4.79	20
	6	体系的手法強調	4.33	4.67	3.83	4.33	4.00	4.16	20
	7	抽象・戦略	4.67	4.67	4.33	5.00	4.50	4.58	20
	8	開発時品質(主特性・副特性)	5.00	4.33	3.67	5.00	4.75	4.42	45
	9	開発時品質(副特性)	5.00	4.33	4.67	5.00	5.00	4.79	10
	10	HAYST法	5.00	4.33	3.83	5.00	4.75	4.47	45

赤文字：指定なし(人力)よりも低い、黒文字：指定なし(人力)と同程度、青文字：指定なし(人力)よりも高い

実験結果①：人力と同等以上

■ 人力と同等～優位：テスト対象、テスト目的、優先度

- テスト対象：人力ではテスト対象外や除外条件などの**テスト要否判断**が欠如
- 優先度：人力や指定なし（AI）では**全く考慮されず**

→人力で不足していた点を、AIでは抽出可能

- テスト目的：**テスト方針に従い、指定なし（AI）よりも具体的な出力**

区分	No	テスト分析方針	テスト対象	テスト目的	テスト条件	優先度	トレーサビリティ・カバレッジ	平均	実行時間(分)
人力	1	指定なし(人力)	3.67	4.67	4.50	1.33	4.50	3.89	360
	2	指定なし(AI)	4.33	3.33	4.33	1.00	3.50	3.47	20
AI	3	ユーザーストーリー	5.00	4.67	4.50	5.00	3.25	4.42	23
	4	リスクベース	5.00	5.00	4.67	5.00	4.50	4.79	26
	5	リスクドリブン	5.00	5.00	4.50	5.00	4.75	4.79	20
	6	体系的手法強調	4.33	4.67	3.83	4.33	4.00	4.16	20
	7	抽象・戦略	4.67	4.67	4.33	5.00	4.50	4.58	20
	8	開発時品質(主特性・副特性)	5.00	4.33	3.67	5.00	4.75	4.42	45
	9	開発時品質(副特性)	5.00	4.33	4.67	5.00	5.00	4.79	10
	10	HAYST法	5.00	4.33	3.83	5.00	4.75	4.47	45

赤文字：指定なし(人力)よりも低い、黒文字：指定なし(人力)と同程度、青文字：指定なし(人力)よりも高い

実験結果②：一部優位/劣位

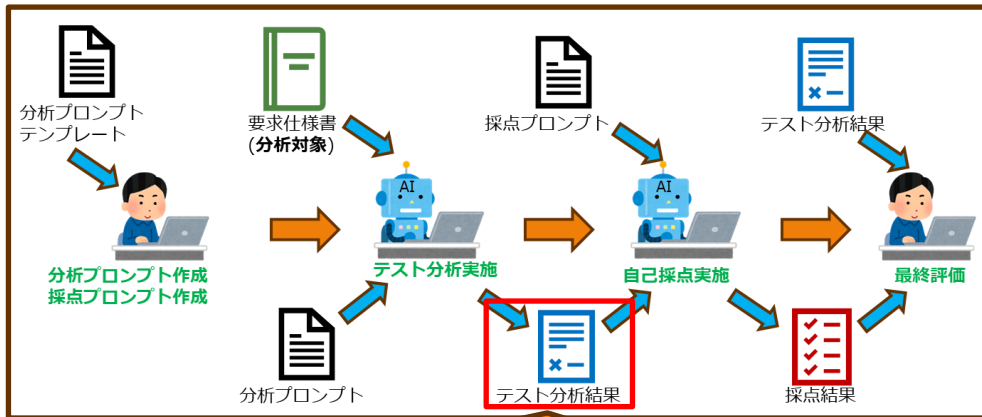
- 一部は優位：トレーサビリティ・カバレッジ
 - ユーザーストーリー方針のみ、人力を下回った
 - テスト分析結果ファイルではチャット内の参照を出力できず
- 【課題】** テスト分析結果ファイルへの参照形式の反映方法を検討

区分	No	テスト分析方針	テスト対象	テスト目的	テスト条件	優先度	トレーサビリティ・カバレッジ	平均	実行時間(分)
人カ	1	指定なし(人カ)	3.67	4.67	4.50	1.33	4.50	3.89	360
	2	指定なし(AI)	4.33	3.33	4.33	1.00	3.50	3.47	20
AI	3	ユーザーストーリー	5.00	4.67	4.50	5.00	3.25	4.42	23
	4	リスクベース	5.00	5.00	4.67	5.00	4.50	4.79	26
	5	リスクドリブン	5.00	5.00	4.50	5.00	4.75	4.79	20
	6	体系的手法強調	4.33	4.67	3.83	4.33	4.00	4.16	20
	7	抽象・戦略	4.67	4.67	4.33	5.00	4.50	4.58	20
	8	開発時品質(主特性・副特性)	5.00	4.33	3.67	5.00	4.75	4.42	45
	9	開発時品質(副特性)	5.00	4.33	4.67	5.00	5.00	4.79	10
	10	HAYST法	5.00	4.33	3.83	5.00	4.75	4.47	45

赤文字：指定なし(人カ)よりも低い、黒文字：指定なし(人カ)と同程度、青文字：指定なし(人カ)よりも高い

実験結果②：一部優位/劣位

- 一部は優位：トレーサビリティ・カバレッジ
 - ユーザーストーリー方針のみ、人力を下回った
 - テスト分析結果ファイルではチャット内の参照を出力できず
- 【課題】** テスト分析結果ファイルへの参照形式の反映方法を検討



提案手法ではテスト分析結果を
中間ファイルとして出力し採点



理由：~~~~~
[要求仕様書4.1 沸騰行為]

想定した出力

```
ハードウェア単体の断線機能 (
* [cite_start]理由：「ソ
明記されており、システムテス
[cite: 193]
```

実際の出カファイルの中身

チャット内でのみ参照できる形式で
テスト分析結果ファイルに出力
→採点時には参照を解決できない。
改善が必要

実験結果③：同等/一部劣位

■ 人間と同等/一部劣位：テスト条件

- 「テスト制約条件」の考慮不足
- 「テスト開始条件」「テスト終了条件」の考慮不足

【課題】 テスト開始・終了・制約条件に対応させるプロンプト修正

区分	No	テスト分析方針	テスト対象	テスト目的	テスト条件	優先度	トレーサビリティ・カバレッジ	平均	実行時間(分)
人カ	1	指定なし(人カ)	3.67	4.67	4.50	1.33	4.50	3.89	360
AI	2	指定なし(AI)	4.33	3.33	4.33	1.00	3.50	3.47	20
	3	ユーザーストーリー	5.00	4.67	4.50	5.00	3.25	4.42	23
	4	リスクベース	5.00	5.00	4.67	5.00	4.50	4.79	26
	5	リスクドリブン	5.00	5.00	4.50	5.00	4.75	4.79	20
	6	体系的手法強調	4.33	4.67	3.83	4.33	4.00	4.16	20
	7	抽象・戦略	4.67	4.67	4.33	5.00	4.50	4.58	20
	8	開発時品質(主特性・副特性)	5.00	4.33	3.67	5.00	4.75	4.42	45
	9	開発時品質(副特性)	5.00	4.33	4.67	5.00	5.00	4.79	10
	10	HAYST法	5.00	4.33	3.83	5.00	4.75	4.47	45

赤文字：指定なし(人カ)よりも低い、黒文字：指定なし(人カ)と同程度、青文字：指定なし(人カ)よりも高い

実験結果③：同等/一部劣位

■ 人間と同等/一部劣位：テスト条件

- 「テスト制約条件」の考慮不足
- 「テスト開始条件」「テスト終了条件」の考慮不足

【課題】 テスト開始・終了・制約条件に対応させるプロンプト修正



人間が想定するテスト条件

- ・ 考慮が必要な環境はハードのみ
- ・ テストの順序性や制約を考慮
- ・ 要員や消化時間は考慮外
(テスト計画で策定すべき)
など…



ギャップ



AIが想定するテスト条件

- ・ 仮想環境やシミュレータも想定
- ・ 順序や制約が曖昧/不足
- ・ リソースとしてテスト要員や
消化時間を考慮
など…

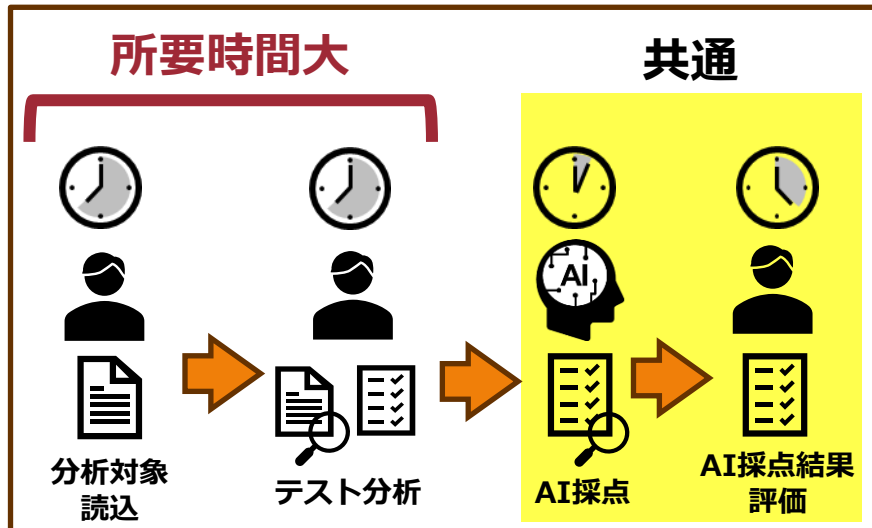
AIと人間で「想定するテスト条件」の**認識差異**がある

→このギャップを埋めるようにプロンプトの検討が必要

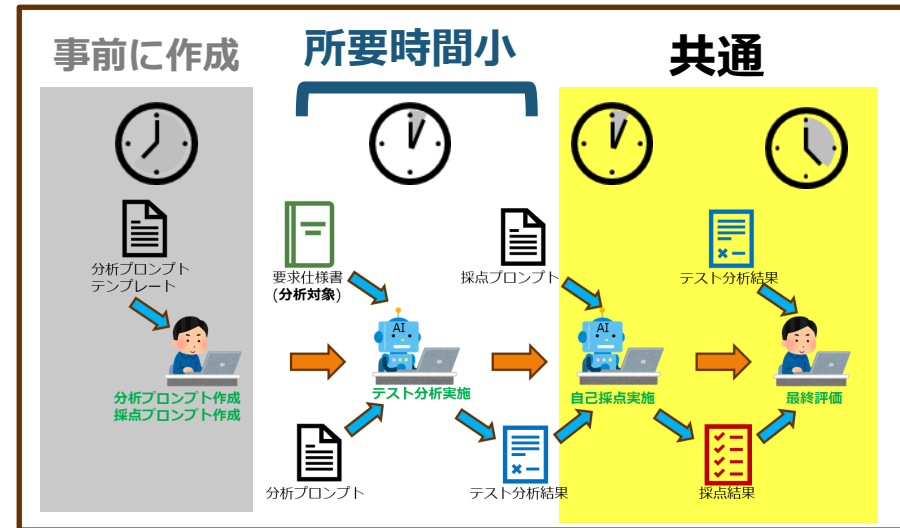
例：「テスト条件では〇〇を評価する」など、厳密に分析方針を指定する、など

実験結果④：実行時間

- 実行時間は大幅に優位な結果に。
 - 人力は 360分 に対して、AIは 平均25分 で完了
 - 人力：分析対象読込～AI採点結果評価完了 までの、テスト分析全体合計時間
 - 提案手法：テスト分析実施～最終評価完了 までの合計時間
 - うち、AI採点とAI採点結果評価での所要時間は共通（同等）
 - 人力では多大な時間がかかる「分析対象読込」と「テスト分析」の分だけ短縮
- ※分析プロンプトは使い回す想定のため、事前に作成とし、実行時間の集計対象外



人間のテスト分析プロセス



提案手法の実行時間

実験結果⑤：その他

■ テスト分析結果/採点結果

- 指定したテスト分析方針通りに分析を実現
- 採点結果から一定の品質での分析を実現

No	テスト分析方針	説明
1	指定なし(人力)	人力（経験30年のQA）によるテスト分析する
2	指定なし(AI)	分析方針を指定せずに「テスト分析を行うこと」のみを記載する
3	ユーザーストーリー	対象製品のユーザーとそのストーリーを想定し、必要なテストを分析する
4	リスクベース	リスクを想定し優先度を設定、優先度ごとにテスト内容を定義する
5	リスクドリブン	リスクを基準にテストの優先順位と範囲を決める分析を行う
6	体系的手法強調	国際規格（ISO/IEC/IEEE 29119）に準拠したテスト分析を行う
7	抽象・戦略	初期段階で決定するテスト戦略立案の観点からテスト分析を行う
8	開発時品質(主特性・副特性)	ISO/IEC 25010の品質特性に基づき、外部品質のテスト分析を行う
9	開発時品質(副特性)	開発時品質の副特性に注目し、ISO/IEC 25010に基づきテスト分析を行う
10	HAYST法	HAYST法で3階層からテスト条件を抽出し、抜け漏れを防ぐ

指定したテスト方針通りに分析

1. 機能テスト (Functional Testing)				
1.1 機能適合性 (Functional Suitability)				
No.	①テスト対象	②テスト目的	③テスト条件・制約	④優先度 (標準)
F-01	沸騰機能 (標準)	水温を沸点(100°C)まで上昇させ、カルク抜きを行う機能が正しく動作することを確認する。	・初期状態：アイドル ・操作：沸騰ボタン押下 ・条件：ヒータONにより水温上昇	5 (基本機能であり全性に因る)
F-02	保温機能 (標準)			
F-03	給湯機能			

1. リスクの特定、分析および評価					
ID	リスク内容	深刻度	発生確率	優先度	分析・評価根拠
R1	過熱による火災・火傷	5	2	10	110°C超過時のヒータ停止失敗は重大な損害に直結するため。
R2	空焚きによる製品故障	4	3	12	全水位センサー等の停止条件が機能しなかった場合、ヒータ稼働を招くため。
R3	温度制御不備 (品質低下)	3	3	9	目標温度 (98, 90, 80°C) の維持失敗は品質低下を招くため。
R4	不適切な状態遷移による誤動作	3	4	12	ボタン操作や蓋開閉に伴う状態不整合を誘発するため。
R5	給湯制御の誤動作	4	2	8	意図しない給湯や停止不可は、高温水リスクがあるため。

2. テスト分析詳細 (優先度順)	
① テスト対象：異常系・安全制御機能	
【トレーサビリティ】要求仕様書 5章：エラー検知 (pot-500-11, pot-500-21, pot-500-31)	
テスト目的: 異常高温やヒータ故障時に、仕様通りの遮断制御と警報が作動し、安全が確保されることを確認する。	
テスト条件:	
高温エラー: 水温110°C超過時に即座にヒータOFF、ブザー30秒鳴動。	
温度下がらずエラー: 保温中、3分以上98°C超過が継続した場合にヒータOFF、ブザー30秒鳴動。	

実験結果まとめ

実験結果

- 人力でのテスト分析と概ね**同等以上の品質**
- 大幅に実行時間が**短縮**できる
- **指定したテスト分析方針**でテスト分析可能
- **一定の品質**でテスト分析可能

課題

- テスト条件、トレーサビリティ・カバレッジなどのテスト分析に対する**プロンプト改修**が必要
- **他のドキュメント**でのテスト分析結果の比較が必要
- 提案手法・分析結果について**人間による評価**が必要

本発表のまとめ

提案手法	<ul style="list-style-type: none">・ 既存のテスト分析手法を「方針」としてAIに与える・ AI分析結果に対してAIによる自己採点する・ AI分析結果・AI自己採点結果に対して人間が評価 これらを一定水準になるまで繰り返す
実験結果	<ul style="list-style-type: none">・ ほとんどの結果で人間と同等以上のテスト分析を実現・ 実行時間の大幅な短縮を実現・ テスト分析方針に従った一定の品質でのテスト分析が可能
今後の課題	<ul style="list-style-type: none">・ プロンプトの改善・ 入カドキュメントのバリエーションによる差異の考察・ 分析結果・採点結果に対する人間の評価が必要

謝辞

本論文の執筆に際し、
以下の方々に丁寧に^ご指導を賜りました。
深く御礼を申し上げます。

- 石川冬樹主査
- 徳本晋副主査
- 栗田太郎アドバイザー

ご清聴ありがとうございました。