



QAエンジニアの暗黙知を可視化する

自動テスト戦略とAIの差分から見たもの

JaSST'26 Niigata 2026.09.18

株式会社LayerX 高橋 諒

高橋 諒 (hashi)

株式会社LayerX バクラク事業部 QA部

経歴

- バルテス株式会社(2019/04-2021/12)
- 個人事業主(2022/01-2022/05)
- 株式会社LayerX(2022/06-)

過去の登壇等

- JaSST'25 Tokyo 事例セッション
[手動 × ローコード × コードベース：3つのアプローチをつなぐ実践](#)
- JaSST'24 Tokyo ミニセッション
[自動テストの成果と失敗談](#)



Agenda

目次

01 なぜやったか

02 何をしたか

03 何が違ったか

04 まとめ

01 なぜやったか

AIコーディングで、機能開発のスピードが変わった

- 1リリースあたりの開発される機能数が増える
- 複雑な機能が短期間でリリースできるようになる

リグレッションテストの範囲が広がり続ける

リグレッションテストは自動テストで守っていく

- 手動テストだけではリリース前のテストが回らない
- 自動テストはAIで実装することが容易になった
- 自動テストを実装するほどメンテナンスコストが増える

自動テストをどう増やすかを考える必要がある

自動テストを増やすときに考えること

- ① 何を自動化する・しないか
- ② レイヤーをどう分けるか
- ③ 優先度をどうつけるか

自動テスト戦略で大きく考える部分になる

今日の問い

これまで

自動化の対象・レイヤー・優先度を、自分で整理して決めてきた

やりたいこと

AIで効率的に自動テスト戦略を組み立てたい

ただし、AIと自分の考えには差分がある

AIと自分のテスト自動化戦略を比較し、差分から暗黙知があるかを明らかにする

02 何をしたか

検証対象：バクラク給与の「随時改定」機能

年の途中で昇給・降給があった従業員の、社会保険の等級を見直す機能

計算

等級を決めるまでの
金額計算と条件判定

状態

従業員別と全体の
確定フローの状態遷移

ファイル

PDFの見た目と
CSVの中身の確認

計算・状態・ファイルが同居していて、どのテストレイヤーで確かめるかが分かれやすい

どうすれば、AIと自分の考えに違いが見えるか

人間とAIがそれぞれ何を根拠に自動テスト戦略の3点を決めているかを知る

1 同じ機能に対して、人間とAIが同じ出力形式の観点表を作成する

2 自動テスト戦略の3点を軸に理由を確認する

実験の条件

AIへの入力

仕様書2ファイル（約1,400行）

実行環境

Claude Code + claude-fable-5.1

回数

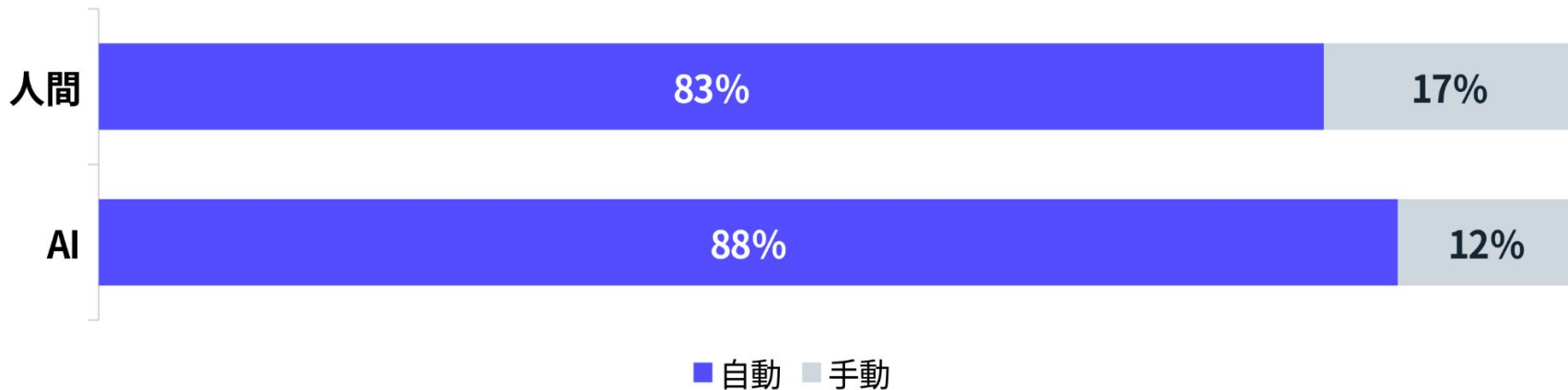
3回

出力形式

観点表（対象／観点／レベル／自動・手動／優先度／理由）

03 何が違ったか

① 自動・手動の割合



AIと比較して、人間の判断は手動テストの割合がやや多い結果となった

人間が手動を選択した理由

ユーザー手動による判定変更

- ・基本的にはユーザーがあまり使わない
想定のため手動テストのみで確認

AIが手動を選択した理由

PDFレイアウト

- ・レイアウトは目視でしか判断できない
ため目視で確認

通勤手当按分

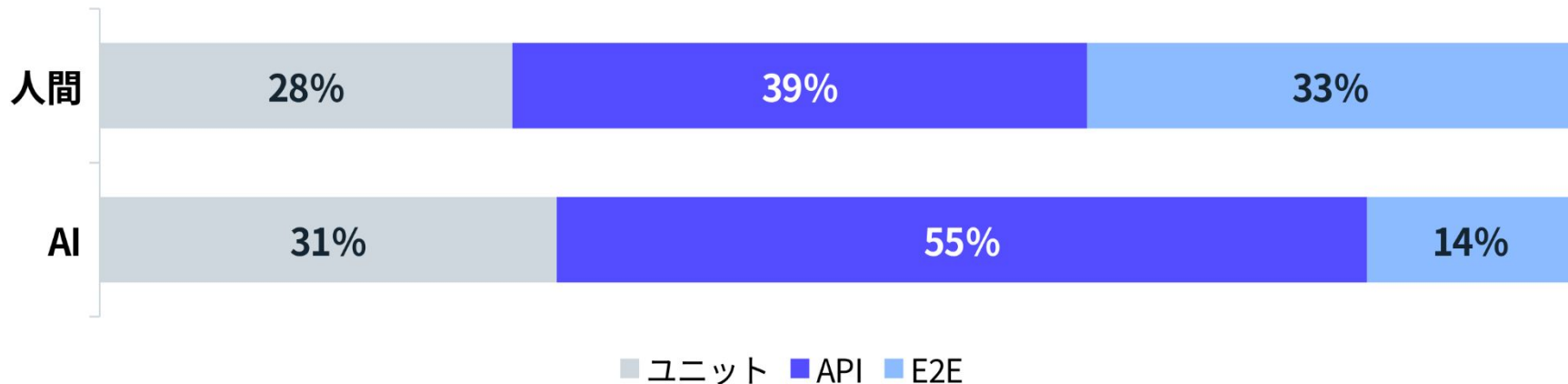
- ・組合せが多く仕様の記述が薄いため、
まず探索的に欠陥を見つける

自分とAIの手動テストとする理由の違い

人間	AI
・利用するケースやシーンが少ない	・目視でないと確認ができない ・仕様の記載が曖昧で自動化できない

AIは人の目が必要なもの、仕様が不明瞭なものを手動テストにするのに対し、
自分は利用するケースの頻度から判断していた

② テストレイヤーの割合



AIと比較して、人間の判断はE2Eテストの割合が多く、APIテストの割合が少ない結果となった

人間がなぜそのテストレイヤーを選択したかの理由

対象者判定

(平均報酬月額計算)

- ・ 法的な計算基準に厳密に従う
必要がある金額計算ロジックのため
ユニットテストで網羅的に行う

最終的な随時改定要否の システム判定

- ・ 全ての計算・判定要素を組み
合わせた総合的なロジック検証
のためAPIテストとする

月次確定・確定解除

- ・ 月次業務の締め処理に該当する
メインのユースケース

AIがなぜそのテストレイヤーを選択したかの理由

等級決定（健保・厚年）

- 等級誤りは保険料誤りに直結するため境界値を網羅するためにユニットテストで確認

標準報酬月額の新計算

- 給与計算確定を解除をトリガーに複数テーブルの更新が発生するため API テストで確認

給与計算への反映

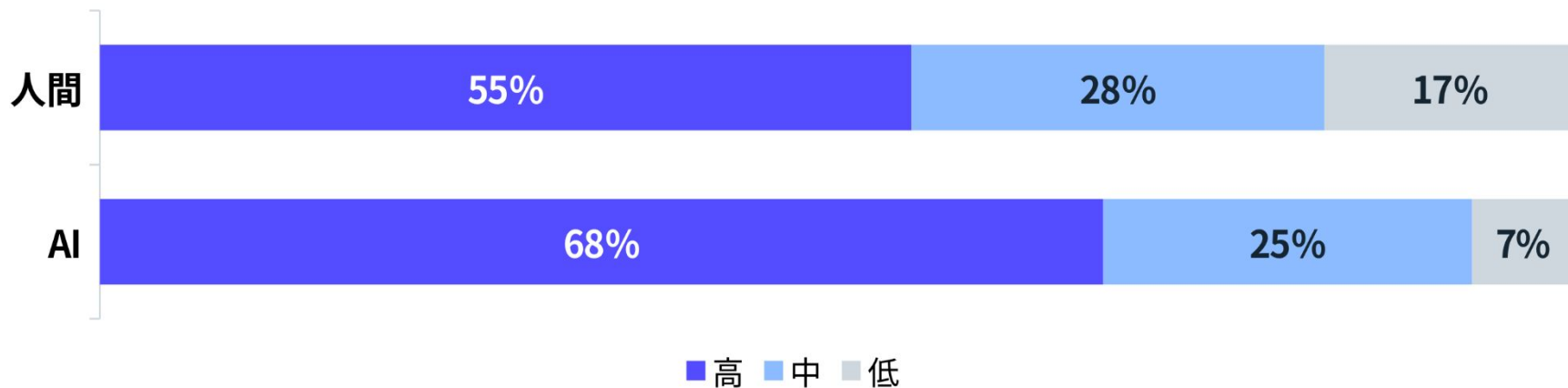
- 随時改定の最終目的であり、給与サービス→HR サービス→給与計算まで跨るユースケースで E2E テストで確認

自分とAIのテストレイヤーを選択する際の理由の違い

	人間	AI
ユニットテスト	・ 答えがひとつに決まる判定・計算	・ 純粋関数である ・ 組み合わせ数の多さ
APIテスト	・ 複合的に組み合わせがある判定・計算	・ DBの状態に依存した
E2Eテスト	・ ユースケースの確認	・ ユースケースの確認

AIは実装を意識した分け方をしているのに対し、
自分は不具合があったときの原因の特定のしやすさを意識していた

③ 優先度の割合



AIと比較して、人間の判断は優先度が高の割合が低く、低の割合が多い結果となった

人間が優先度を低くした理由

フィルタ

- ・業務を止めるほどのリスクはない
- ・ブラウザの標準的な検索機能や出力したCSVデータからエクセル等で代替が可能なため

CSVインポート

- ・手間はかかるが業務を止めるほどのリスクはない
- ・基本的にはシステムの自動判定がメイン
- ・毎月多くの従業員を対象として行われる業務ではなくCSVインポートではなく画面からの変更も可能であるため

AIが優先度を低くした理由

メモ欄の更新

- ・金額に影響しない

届出の備考欄自動入力

- ・特定の条件のみの問題
- ・届出を出力すると備考が空になるが
手入力回避可能なため

自分とAIの優先度を下げている理由の違い

人間	AI
・金額に影響しない ・ほかに代替できる手段がある ・利用するケースやシーンが少ない ・業務が止まるほどの影響がない	・金額に影響しない ・ほかに代替できる手段がある

AIには金額への影響、代替手段の有無が理由になっており、
自分はとして機能の使われ方と業務への影響から優先度を判断していた

04 まとめ

自分の暗黙知ってなんだったんだろう

① 自動化しなかった理由

「使うユースケースが
少ないため」

② テストレイヤーを分ける理由

「不具合があったときの原因の
特定のしやすさ」

③ 優先度を低くした理由

「業務が止まるほどの
影響がない」



テストを書いた後にどのくらい価値があるかを判断していた

まとめ

- ・人間とAIで自動テスト戦略を比較して3つの判断軸について理由の比較を行った
- ・自分の場合は「テストを書いた後の価値」を暗黙的に重要視していた
 - ・今後発生する、第三者への3つの判断軸による意思決定の背景を説明する際に、「利用頻度」と「業務への影響」に基づいていることを、明に伝えるようにしていきたい

ご静聴ありがとうございました