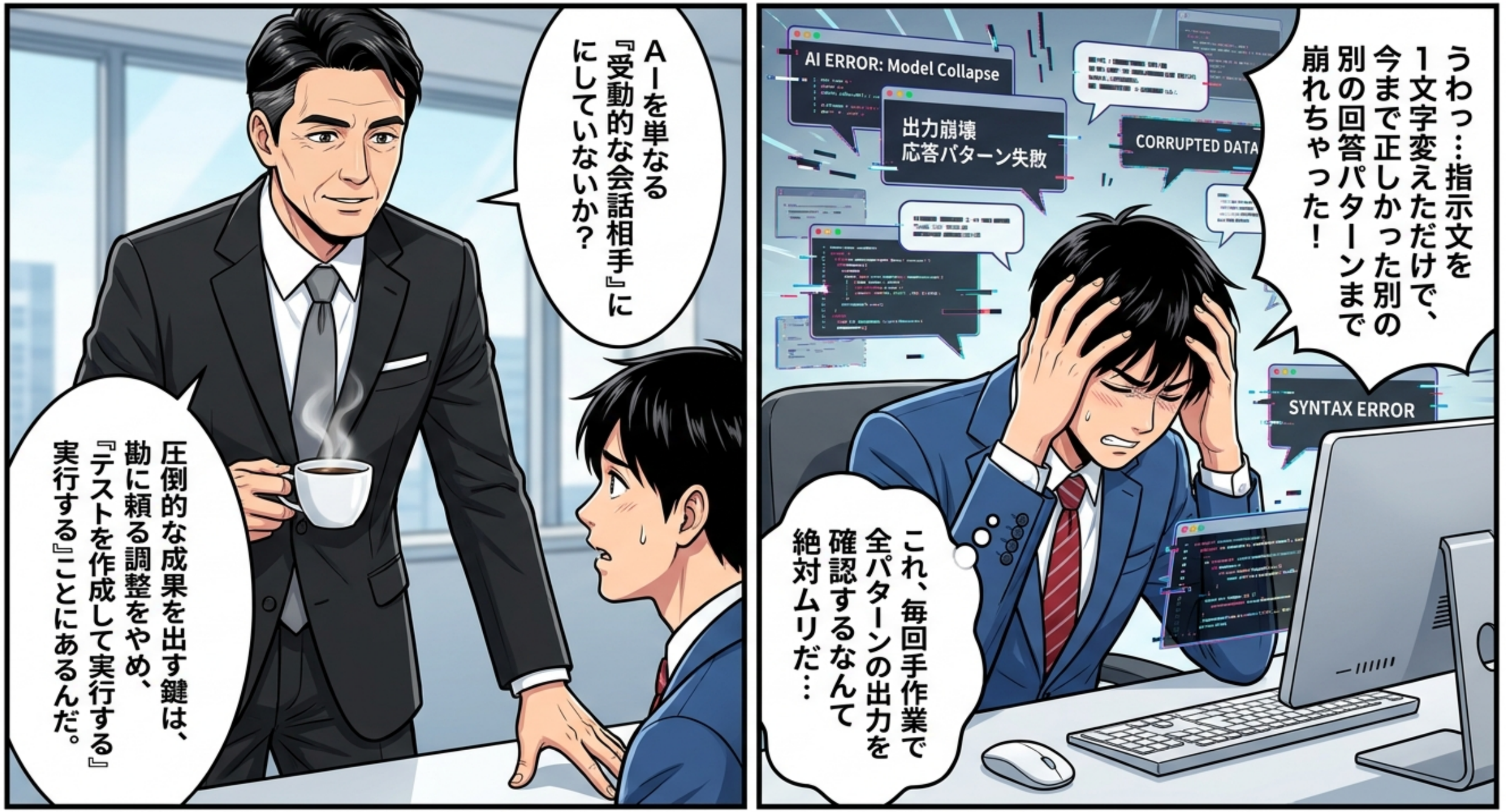
Claudeでマスターする「テストの作成と実行」

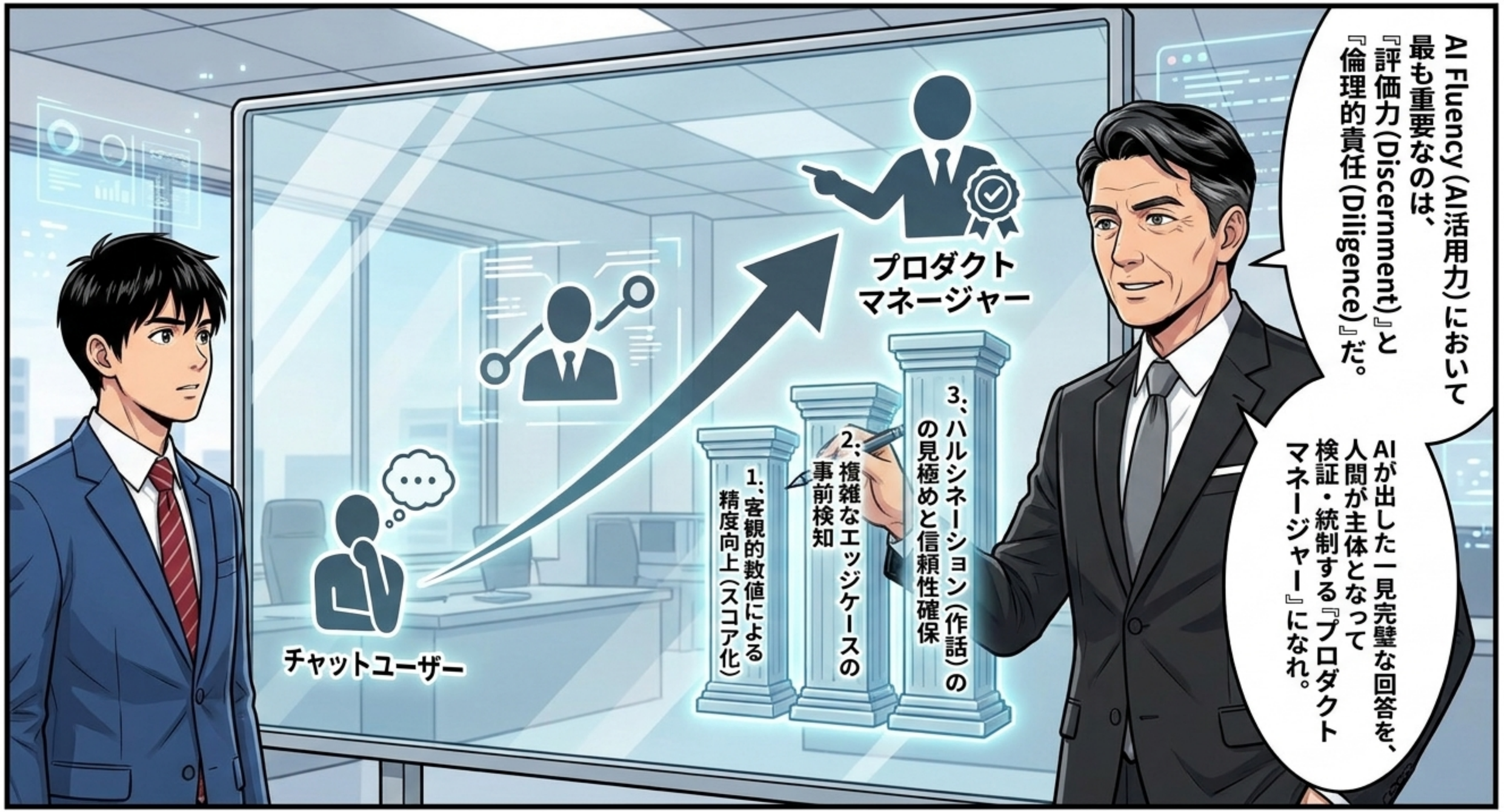
AIの精度を高め、作業の信頼性を劇的に引き上げるステップバイステップ手順書



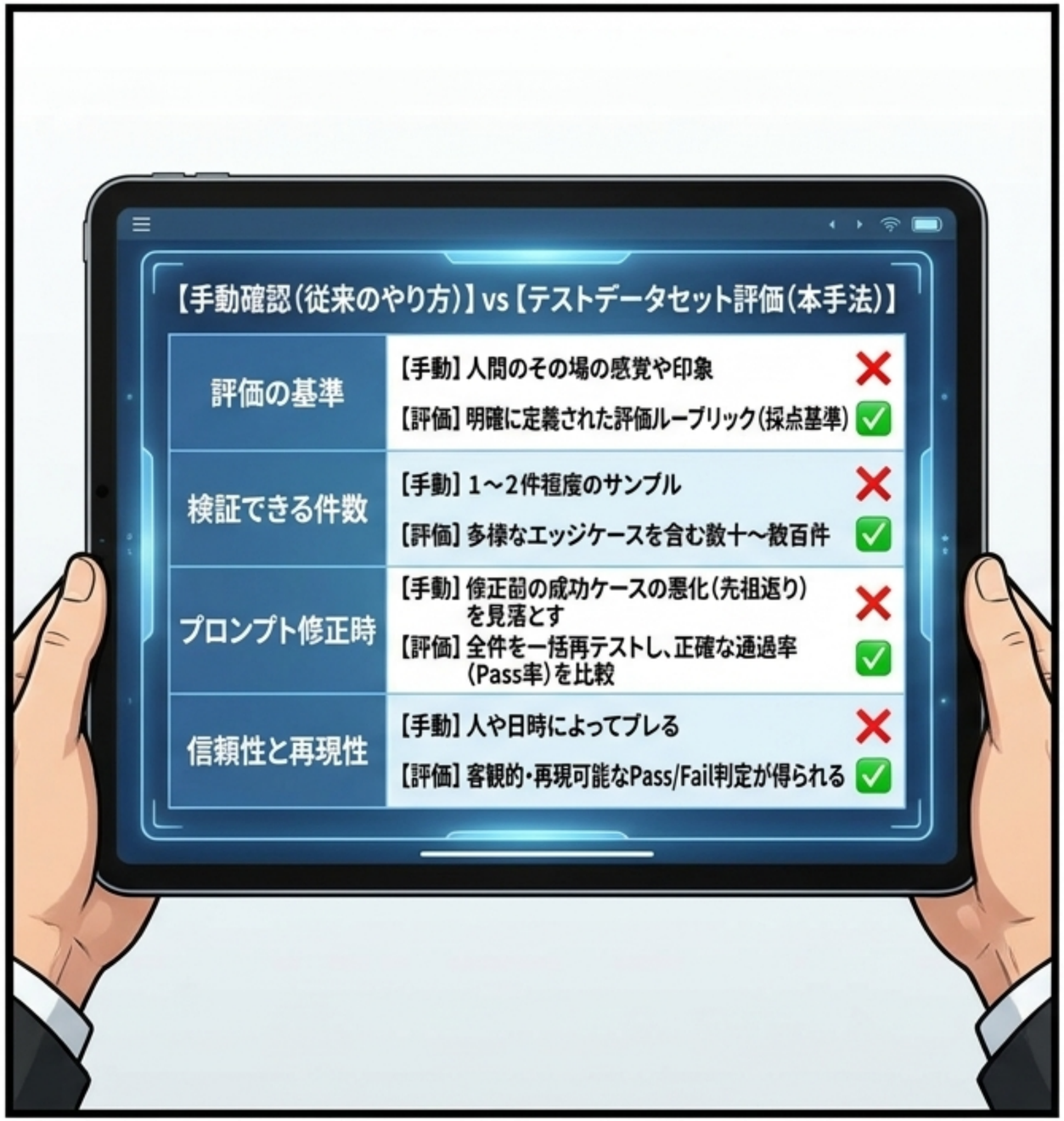
Gemini Notebook



Gemini Notebook

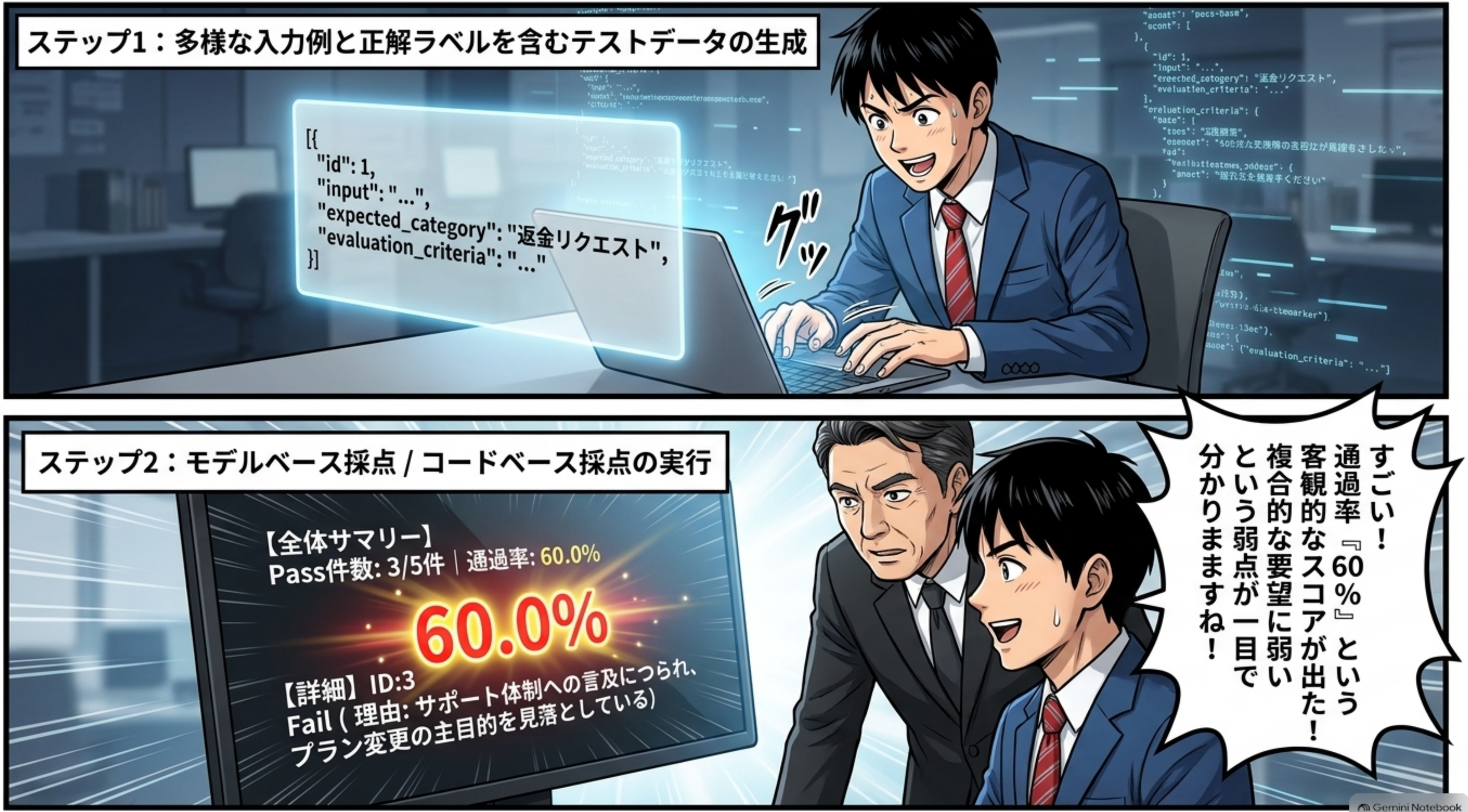


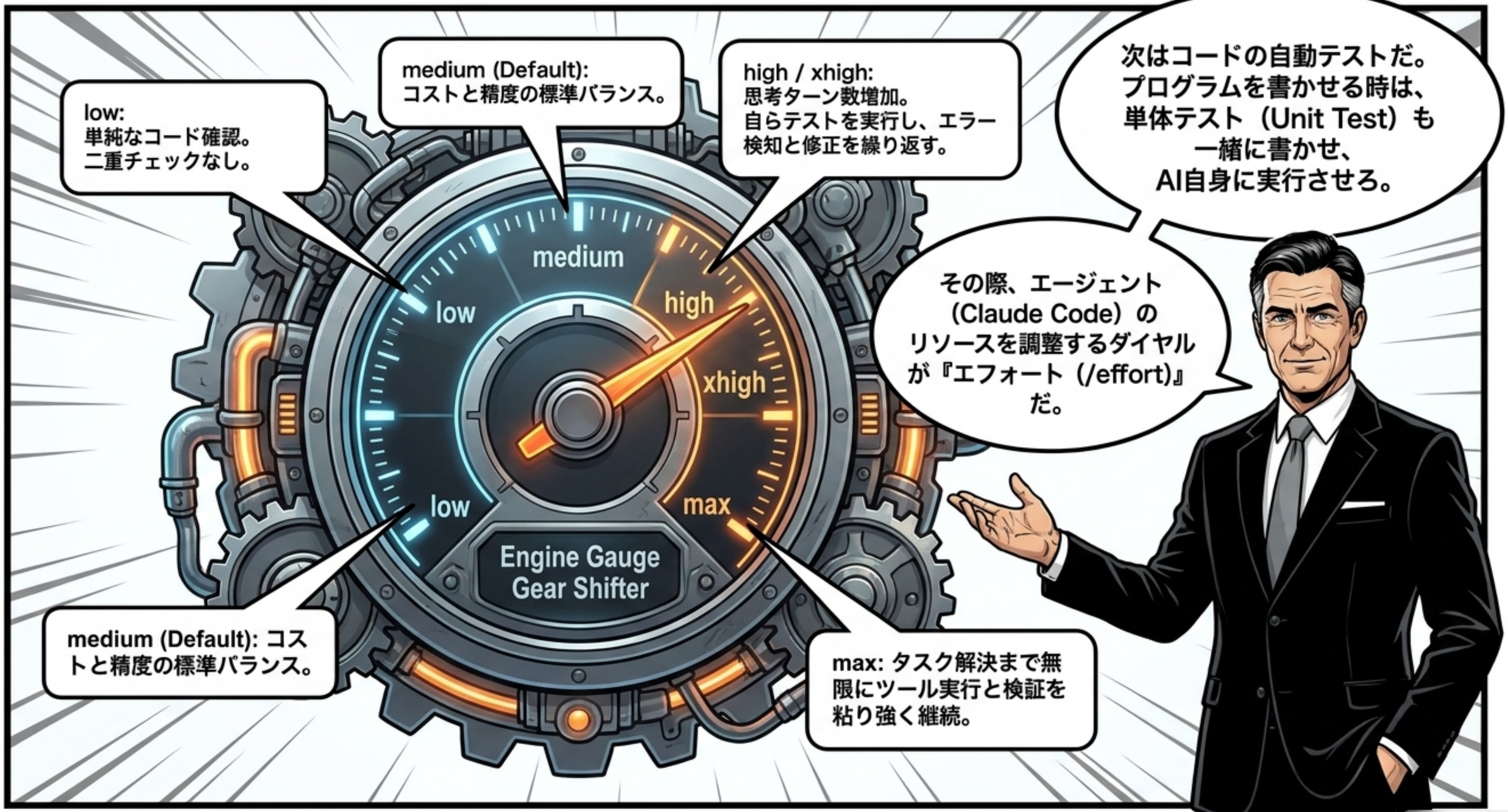
Gemini Notebook





Gemini Notebook


ステップ1：多様な入力例と正解ラベルを含むテストデータの生成
[{
"id": 1,
"input": "...",
"expected_category": "返金リクエスト",
"evaluation_criteria": "..."
}]
グッ
ステップ2：モデルベース採点 / コードベース採点の実行
【全体サマリー】
Pass件数: 3/5件 | 通過率: 60.0%
60.0%
【詳細】ID:3
Fail (理由: サポート体制への言及につられ、
プラン変更の主目的を見落としている)
すごい！
通過率『60%』という
客観的なスコアが出た！
複合的な要望に弱い
という弱点が一目で
分かりますね！


low:
単純なコード確認。
二重チェックなし。
medium (Default):
コストと精度の標準バランス。
high / xhigh:
思考ターン数増加。
自らテストを実行し、エラー
検知と修正を繰り返す。
次はコードの自動テストだ。
プログラムを書かせる時は、
単体テスト（Unit Test）も
一緒に書かせ、
AI自身に実行させろ。
その際、エージェント
（Claude Code）の
リソースを調整するダイヤル
が『エフォート（/effort)』
だ。
low
medium
high
xhigh
max
Engine Gauge
Gear Shifter
medium (Default): コス
トと精度の標準バランス。
max: タスク解決まで無
限にツール実行と検証を
粘り強く継続。

```
❯ claude --effort high
● Running tests: pytest test_csv_to_json.py
FAILURES: AssertionError: Expected 'start_date' key ❌
⚠️ Test failed. Analyzing cause with increased reasoning effort...
→ Auto-fixing csv_to_json.py... DONE ⚙️
● Re-running tests... PASSED successfully! ✅
```



Gemini Notebook

トピック3：AI出力の嘘を見破る心理的盲点

⚠【アーティファクト効果】
研究データでも、美しい文書が出力されると人間の「事実確認」が-3.7%、「推論への疑問」が-3.1%低下することが実証されている。

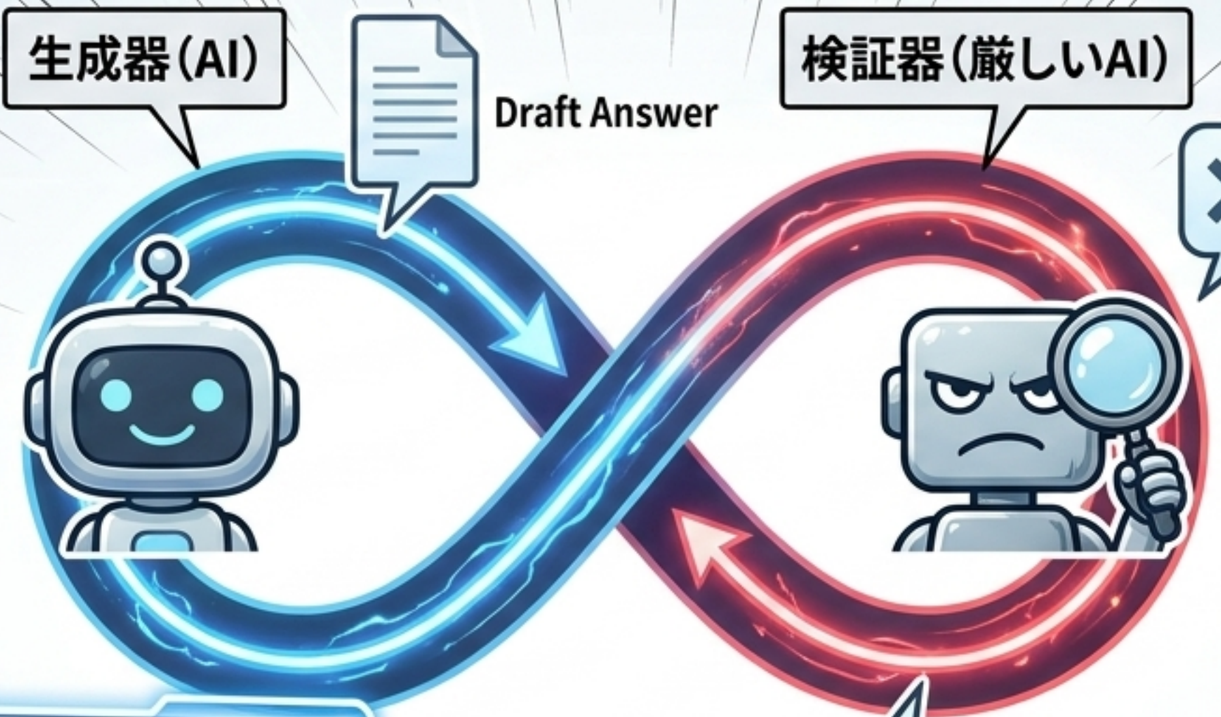


Gemini Notebook

リスクに対抗するための最強の盾が、『生成器と検証器のループ（Generator-Validator Loop）』だ。

AIに回答を出させた後、あえて『厳格な検証者』のペルソナを与え、別のプロンプトで再点検させるんだ。

AIに回答を出させた後、あえて『厳格な検証者』のペルソナを与え、再点検させるんだ。





【検証テストでチェックすべき3大要素】

1. 事実関係（ファクトチェック）：
   固有名詞や数値が一次資料と一致するか？
2. 推論の飛躍：
   前提から結論への論理の欠落（BやCのすっ飛ばし）はないか？
3. 欠落しているコンテキスト：
   組織独自の制約や潜在リスクが見落とされていないか？

Gemini Notebook





Gemini Notebook



Gemini Notebook



今日私が作成した以下の【テキスト/アイデア】の弱点を1分でテスト検出してください。

Copy

【テスト指示】：
1. 論理の飛躍や説明が不足している「ギャップ」を1点指摘してください。
2. 懐疑的な第三者が提出しそうな突っ込みや反論を1点挙げてください。
3. 信頼性を高めるために、今すぐ追加・修正すべき具体的アクションを1点教えてください。前置きは不要です。

Gemini Notebook