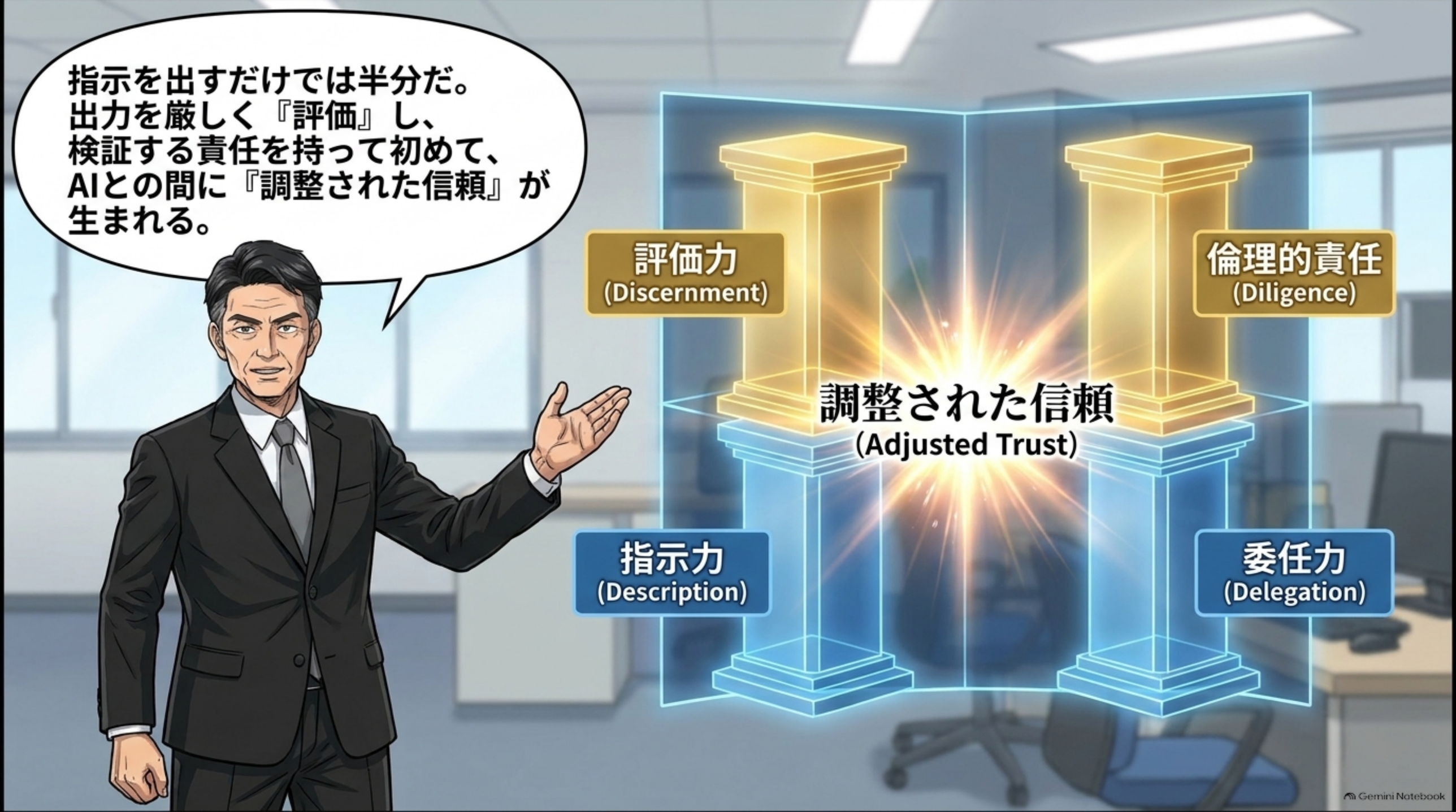
【完全ガイド】Claudeでマスターする「テストの作成と実行」

AIの出力を鵜呑みにするな!
品質を劇的に高める評価ワークフロー



Gemini Notebook


プロンプトを一回
入力するだけで、
完璧な回答が
出ました！
完璧なるけで
完璧な回答が
出出ました！
単発の成功に満足してはい
条件が少し変わっただけで、
予期せぬエラーが起きるぞ。
起きるぞ。
AIへの「盲信」から抜け出し、
「評価者」へ変わる時が来た。
Gemini Notebook


指示を出すだけでは半分だ。
出力を厳しく『評価』し、
検証する責任を持って初めて、
AIとの間に『調整された信頼』が
生まれる。
評価力
(Discernment)
倫理的責任
(Diligence)
調整された信頼
(Adjusted Trust)
指示力
(Description)
委任力
(Delegation)
Gemini Notebook

このクレームメール1件でテストしたところ、正しく分類されました！

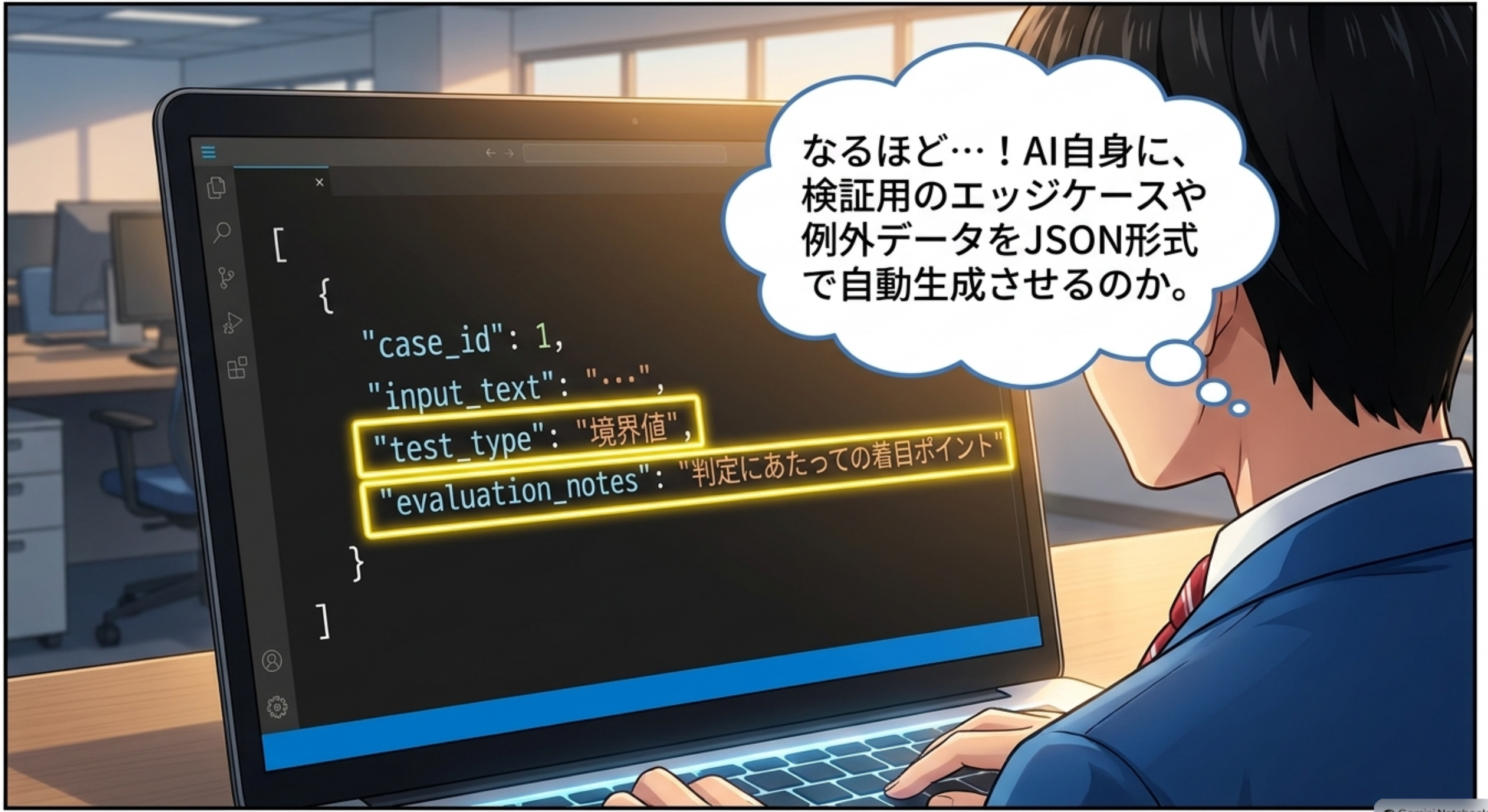
テストデータセット



1件だけでは不十分だ。本番環境の多様な条件を網羅した『テストデータセット』で一元評価する必要がある。

1. 正常系 (Normal)：
   明確な製品質問
2. 境界値 (Edge Cases)：
   怒りを含む質問、複数要素の混在
3. 例外系 (Exceptions)：
   支離滅裂な文章

Gemini Notebook

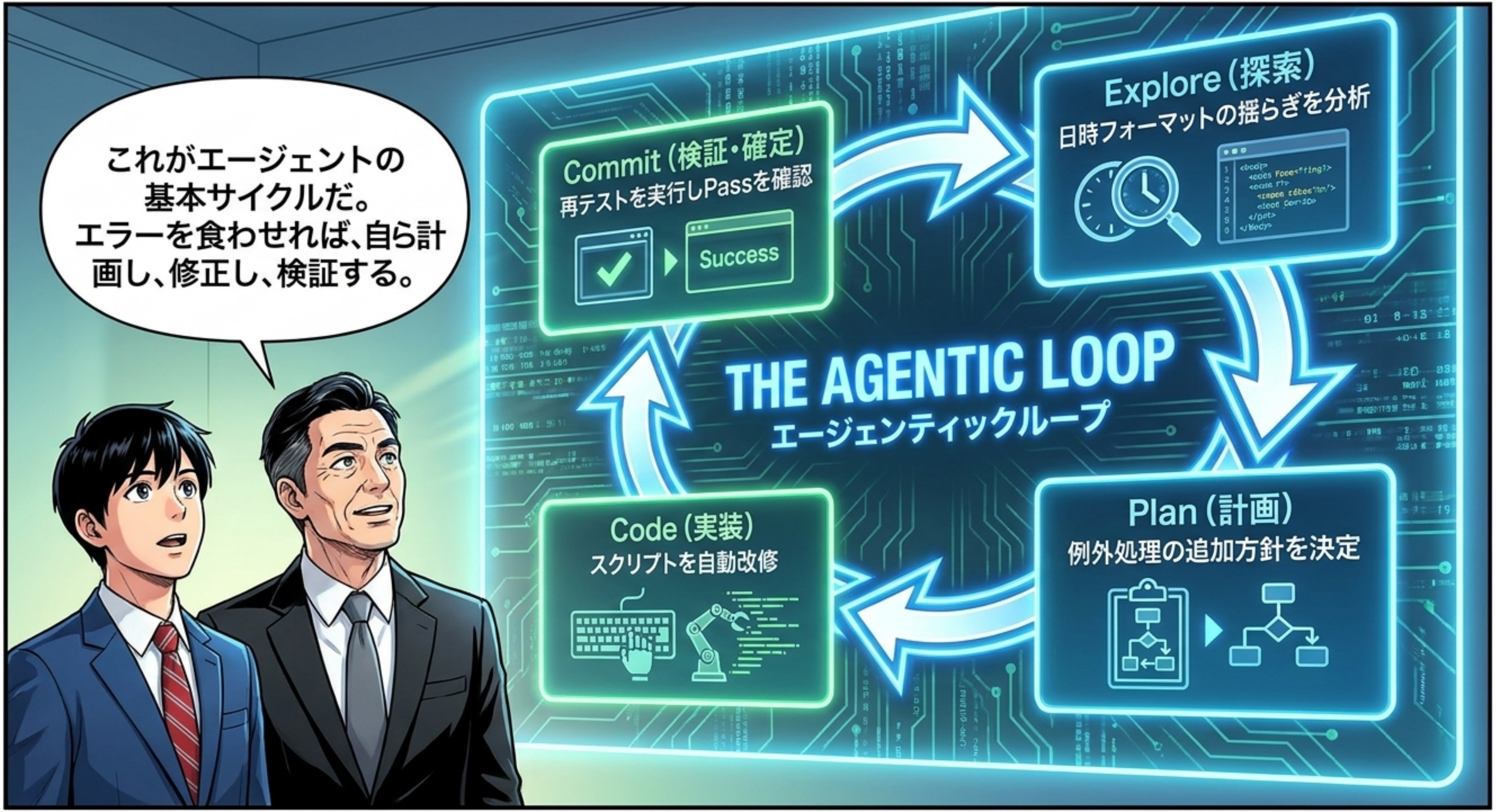


Gemini Notebook


【モデルベース採点】
Claudeの高い推論能力を活用。定性的なニュアンス、トーン、複雑な分類の妥当性を判定。
【コードベース採点】
Pythonや正規表現による機械的判定。完全一致やJSON構造の整合性など、定量的な厳密検証。
目的に応じて使い分けろ。ニュアンスはモデルに、厳密なフォーマットはコードに採点させるんだ。


厳格な品質保証
アナリスト
AIの『おもねり現象
(Sycophancy)』に騙されるな！
ユーザーに同調して
甘い採点をする傾向がある。
素晴らしい
プロンプトです！
完璧です！
制約プロンプト：「一切のおもねりを排除し、
厳しい異議・疑義や改善点を率直に指摘してください」

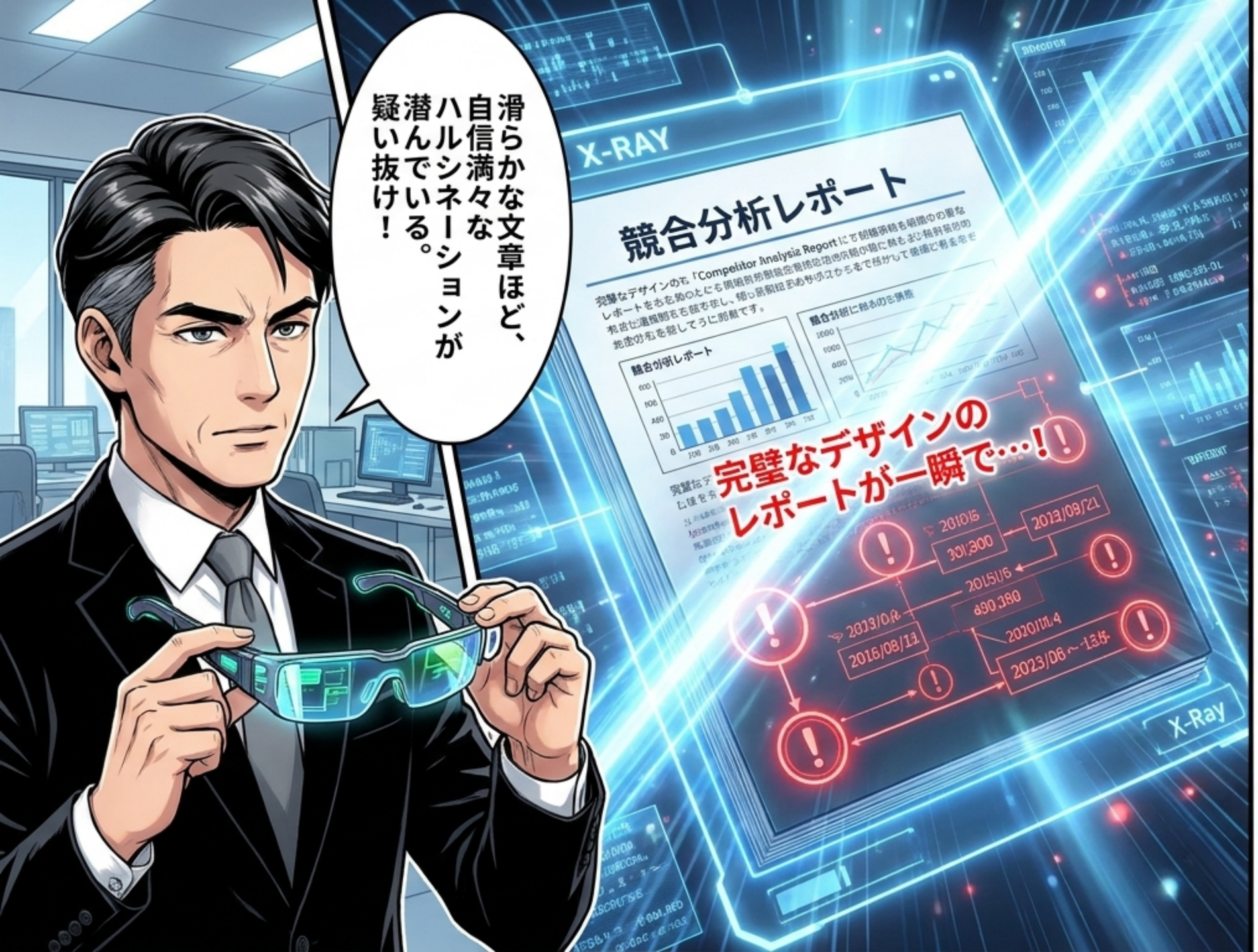

小トピック2：コードの確実な動作検証
焦るな。
エラーログも
Claude Codeに
解析させるんだ。
Error Message Box
Traceback: ValueError:
time data '2025/01/27'
does not match
format '%Y-%m-%d'
うわっ！
日付のフォーマット違いで
エラーが…！
Gemini Notebook



Gemini Notebook


medium/auto:
コストと精度の
最適バランス
high:
思考とツール呼び出し
を反復し正確性UP
low:
高速・自己チェックなし
max:
セッション限定の
徹底的な二重チェック
未知のライブラリや複雑なバグに直面したら、
『エフォートレベル』のダイヤルを上げろ。
AIの"働く執念"が変わる。
Gemini Notebook


すごい…
完璧なデザインの
レポートが一瞬で…！
競合分析レポート
見た目の美しさに！
それが『アーティファクト
効果』の罠だ。
研究データ：整った成果物を見ると、
人間の「事実確認 (Fact-Checking)」
行動は -3.7pp 低下する。
Gemini Notebook



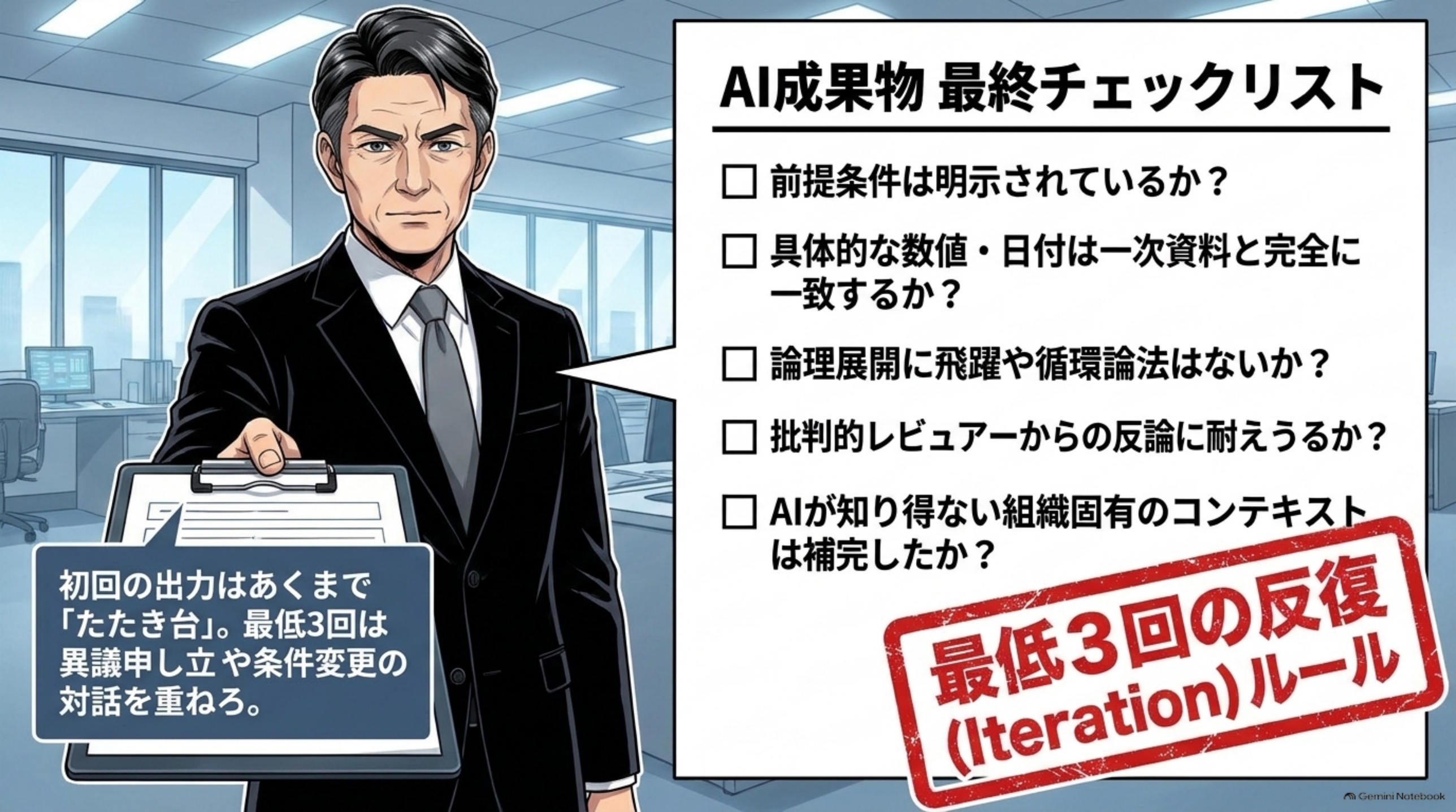
クリティカル・ストレステストの3眼

1. 前提条件の抽出: 結論の裏にある暗黙の前提を暴く

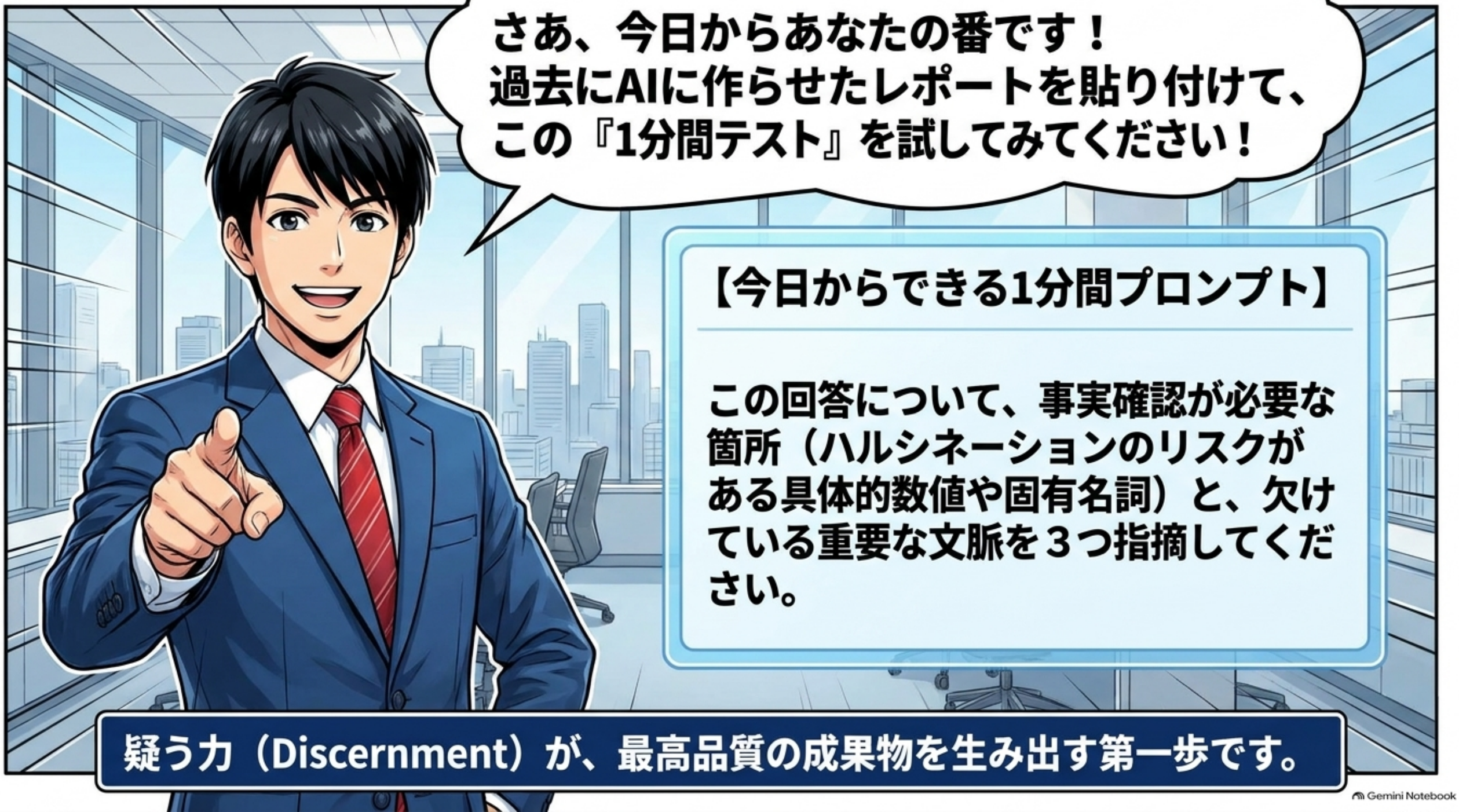
2. 反事実的質問: 「競合の視点から反論せよ」と揺さぶる

3. 一次情報源のクロスチェック：固有名詞や数値の作話(ハルシネーション)を疑う

Gemini Notebook


AI成果物 最終チェックリスト
□ 前提条件は明示されているか？
□ 具体的な数値・日付は一次資料と完全に一致するか？
□ 論理展開に飛躍や循環論法はないか？
□ 批判的レビュアーからの反論に耐えうるか？
□ AIが知り得ない組織固有のコンテキストは補完したか？
初回の出力はあくまで「たたき台」。最低3回は異議申し立や条件変更の対話を重ねろ。
最低3回の反復(Iteration)ルール
Gemini Notebook


もうAIの回答をただ受け取るだけの『受け身のユーザー』ではありません。私自身が品質をコントロールします。
その通りだ。評価し、検証する力こそが、これからの時代における君の最強の武器になる。
プロンプトエンジニアリングから、「評価ワークフロー」へ。
確信を持ったパートナーへの進化。
Gemini Notebook



Gemini Notebook